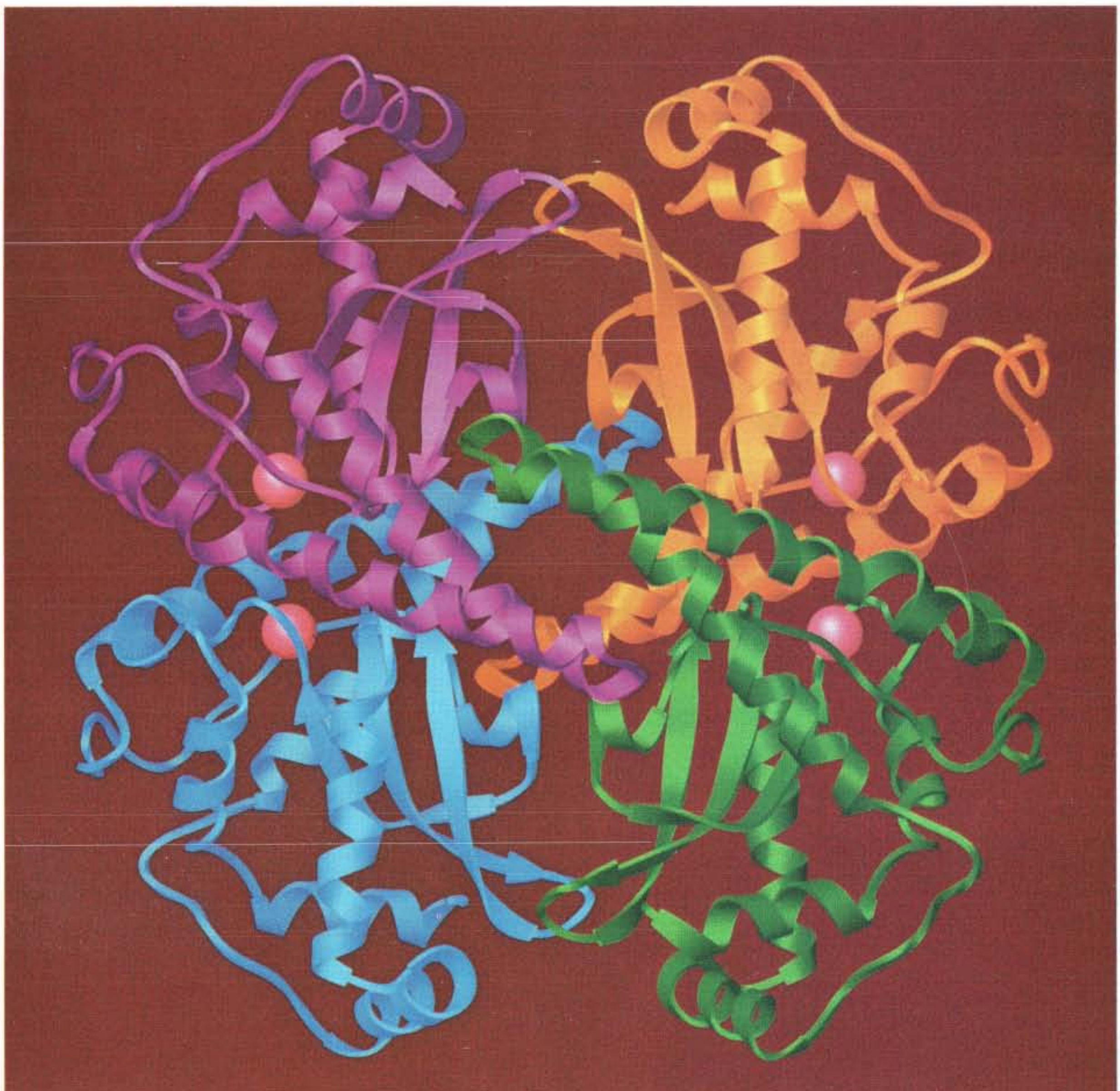
SCIENTIFIC AMERICAN

DECEMBER 1992

U.K. £2.50 $3.95

SPECIAL SECTION
NEW CHALLENGES FOR 1993

Toward atoms made of antimatter.
Economics as an experimental science.
The simple lesson of Asian education.



Radical interceptor: *this antioxidant enzyme may be one of the body's defenses against aging.*

WHEN WAS THE LAST TIME YOU HIT AN ELK

Sweden is a land populated by many elks that sometimes leave their forests and stray onto unprotected roads.

The adult elk is large, heavy and mostly dark. In winter, the Swedish countryside is mostly dark as well. Which explains why surprised Swedish

drivers and elks often collide. Apart from elks, Sweden offers other unexpected road hazards like ice, snow and mud.

Fortunately, Swedish roads are populated with many Saab 9000s. (In four separate international car safety studies, Saab headed the lists.) Saabs

ıave crumple zones at both ends :o absorb the energy of a collision ;hould you unexpectedly make contact vith a large, dark animal.

And airbags* and seat-belt tensio- ıers should you be thrown forward. 3ut if you take prompt avoiding action, :he Saab 9000 has ABS** fitted as standard to prevent your wheels from locking when you swerve while braking heavily.

The elk might be left undamaged but bemused. And wondering why every intelligent human doesn't drive a Saab 9000.

SAAB. TRULY SWEDISH.

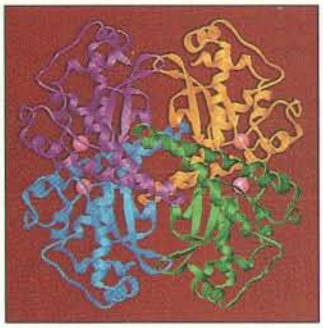
Saab 9000 Model Year '93. Standard in certain countries. Available as extra everywhere.
*Non standard on all Saab 900 models in DK, IS, SF, N.

SAAB

SCIENTIFIC AMERICAN

December 1992 Volume 267 Number 6

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 1992 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111, or fax: (212) 355-0408.



THE COVER image shows the newly deciphered, four-part structure of the human antioxidant enzyme manganese superoxide dismutase, the form of superoxide dismutase active in the mitochondria of cells. Despite such defenses, oxidative damage can accumulate in mitochondria and may thus play a role in aging and in age-related disorders (see "Why Do We Age?" by Ricki L. Rusting, page 86). Gloria E. O. Borgstahl, Hans E. Parge, Michael J. Hickey, Robert A. Hallewell and John A. Tainer of the Scripps Research Institute solved the structure.

THE ILLUSTRATIONS

Cover image by John A. Tainer, Gloria E. O. Borgstahl, Alexandre K. Shah and Michael E. Pique, Scripps Research Institute

Page	Source	Page	Source
33	Eiji Miyazawa/Black Star	75	Steve Dreiseszun/ Viewpoint Photographers
34	Harold W. Stevenson (*photograph*); Laurie Grace	76	Johnny Johnson (*top*), Sygma (*bottom*)
35	Tony Freeman/Photo Edit (*photograph*); Laurie Grace	77	P. Vauthey/Sygma
36–37	Laurie Grace	79	Dean Lee/The Wildlife Collection
38	David Young-Wolff/Photo Edit (*photograph*); Laurie Grace	80	Patricia J. Wynne
		82–83	Patricia J. Wynne
41	George Retseck	84	Dean Lee/The Wildlife Collection (*left*), Robert M. Seyfarth and Dorothy L. Cheney (*right*)
43–44	George Retseck		
45	Gerald Gabrielse		
46–47	George Retseck	86–87	Robert Prochnow
49	Dan Wagner	88–89	James R. Smith, Baylor College of Medicine (*top*), Johnny Johnson (*bottom*)
50–55	Ian Worpole		
56–58	Courtesy of the New York Public Library	90	Joseph L. Graves, University of California, Irvine (*left*), Stephanie Rausser (*right*)
59	Topkapi Saray Museum, Istanbul	91	Ken Abbott, University of Colorado (*left*), Sinclair Stammers/Science Photo Library/Photo Researchers, Inc. (*right*)
60–61	Courtesy of the New York Public Library		
62	Theodore de Bry; courtesy of the New York Public Library	92	Tomo Narashima
65	Ross Marino/Retna Ltd.	93	Tomo Narashima (*left*), Johnny Johnson (*right*)
66–69	Carol Donner	94	Cindy L. Fisher, Michael E. Pique and Elizabeth D. Getzoff, Scripps Research Institute
70–71	Robert T. Sataloff		
72–73	J. P. Laffont/Sygma		
74	Johnny Johnson	96–98	Michael Field and Martin Golubitsky

SCIENTIFIC AMERICAN®

Established 1845

LETTERS TO THE EDITORS

Using Fetal Tissue

I read "Aborting Research," by Tim Beardsley ["Science and the Citizen," SCIENTIFIC AMERICAN, August] with some dismay, not so much because of what was said as because of what was unsaid. He states that there is a "clear consensus" of researchers who believe that results of fetal tissue transplants in Parkinson's disease sufferers "are encouraging." Yet C. G. Clough, a British researcher, having evidently missed this consensus, wrote in the June 1, 1991, issue of *Lancet*, "Although about one hundred operations with fetal implants have now been completed, there is little evidence of implant survival.... The technical difficulties of the procedure suggest that neural implantation is unlikely to benefit many patients with Parkinson's disease."

According to William Landau, the chief of neurology and neurosurgery at Washington University, physicians have known since 1909 that making lesions into the brains of Parkinson's patients has caused improvement in their symptoms. In 1990 Landau wrote that the improvements after transplants were not necessarily caused by the fetal tissue but rather were "compatible with operative trauma [and] post-operative tissue reaction."

ROBERT A. MCFADDEN
Ad Hoc Committee in Defense of Life
Washington, D.C.

Quantum Quandary

"Quantum Philosophy," by John Horgan [SCIENTIFIC AMERICAN, July] contains some inaccuracies in its description of the work of GianCarlo Ghirardi, Alberto Rimini, Tullio Weber and myself. We alter Schrödinger's equation so that the wave function can describe the reality we see around us rather than just the probability of that reality, as the standard quantum theory does.

Our proposal is testable, but it has not yet been tested, much less refuted as the article states. The beautiful two-slit neutron interference pattern obtained by Anton Zeilinger, Roland Gähler and Anthony G. Klein is compatible with the predictions of both the standard quantum theory and our stochastic modification of it.

As for the quotation implying that Zeilinger regards our theory as a "dead end," in private correspondence he has written that his opinion "is certainly more subtle. I think your approach is important from a basic point of view because that possibility has to be tried out, but I personally believe that the Copenhagen interpretation will survive in the end." For another opinion, consider that of John Bell, who in 1990 described the stochastic modification of quantum mechanics as the most important new idea about the foundations of the field to emerge during his professional lifetime.

PHILIP PEARLE
Department of Physics
Hamilton College

In his article, John Horgan discussed a hypothetical experiment to pass an amoeba through an interferometer to test concepts of quantum mechanics. He then wrote, "Getting a slightly larger and more intelligent organism, for instance, a philosopher, to take two paths through a two-slit apparatus would be even trickier."

One would like some evidence that philosophers are more intelligent than amoebas at all. That minor complaint about an otherwise excellent article does not affect the value of the experiment. It's worth a try.

LEE HARDING
Coquitlam, British Columbia

Applied Physics Only

Roy F. Schwitters wants the taxpayers of America to cough up at least $6 billion for the Superconducting Super Collider ["The Frustrations of a Quark Hunter," by Russell Ruthen; "Science and the Citizen," SCIENTIFIC AMERICAN, September]. Yet not one word in the profile is said to justify this enormous expense other than Schwitters's desire to be the Captain Ahab of the top quark. This conceit is a beautiful illustration of the arrogance of today's theoretical scientists.

I suggest that Schwitters and his colleagues climb down from their ivory towers and join the fight for the future of planet Earth. After all, there will be no more theoretical science if humanity sinks into oblivion because of severe overpopulation.

WARREN C. LITSINGER
Newport, Ore.

A Bowdlerized Botanica

Carolus Linnaeus scandalized 18th-century Europe when he chose to classify plants by their sexual parts. The illustration on page 20 of "How Many Species Inhabit the Earth?" by Robert M. May [SCIENTIFIC AMERICAN, October], offers an example of his "Methodus Sexualis": it depicts the stamens of male flowers in varying arrangements. The caption emasculates Linnaeus's work by saying that he "organized plants based on leaf shapes."

ANSELM ATKINS
Decatur, Ga.

Explosive Results

In "Biological Roles of Nitric Oxide" [SCIENTIFIC AMERICAN, May], Solomon H. Snyder and David S. Bredt mistakenly credit the invention of nitroglycerin to Alfred Nobel. Ascanio Sobrero first prepared nitroglycerin in 1846 or early 1847. Nobel's claim to fame is for taming the highly sensitive liquid explosive by diluting it with inert materials (such as sawdust or diatomaceous earth) to make dynamite.

DARREN NAUD
Center for Explosives
Technology Research
Socorro, N.M.

ERRATUM

The caption on page 65 of "Diamond Film Semiconductors" [October] incorrectly describes the diamond lattice. The carbon atoms are arranged in a cubic lattice of hexagonal rings. One such ring is shown in red.

Because of the volume of mail, letters to the editor cannot be acknowledged. Letters selected for publication may be edited for length and clarity. Unsolicited manuscripts must be accompanied by a stamped, self-addressed envelope.

50 AND 100 YEARS AGO

DECEMBER 1942

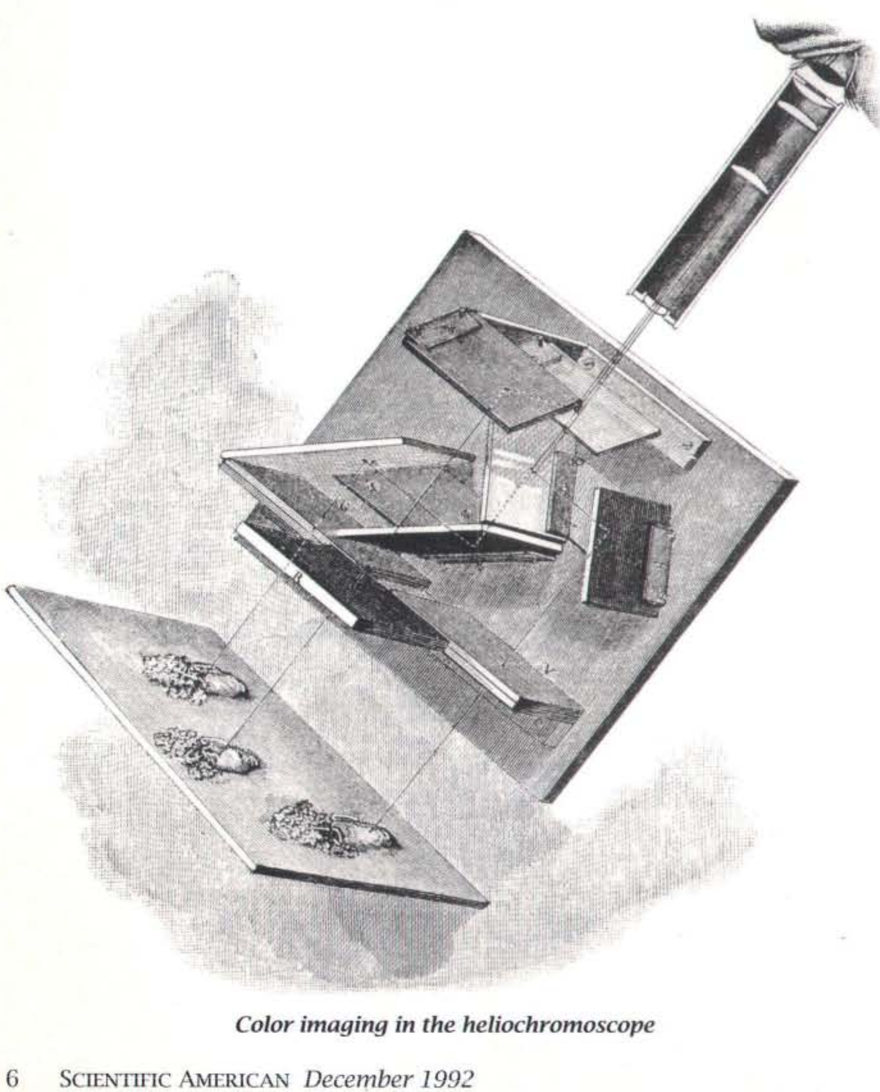
"We have as yet no idea at all why or how so much of the matter in the galaxy ever came to segregate into huge lumps like the stars. It is in fact hard to see how this could happen, if we start with the conventional assumption that the matter was 'originally' distributed with rough uniformity through galactic space. On this basis, the stuff that ultimately formed a star would have been distributed through a volume at least a light-year in diameter. Unless the relative motions within this scattered assemblage were exceedingly small, the mass, when it shrank to the dimensions of a star, would be rotating far faster than the average star."

"The first discovery, in 1927 at Folsom, New Mexico, of peculiarly shaped implements of human manufacture associated with the bones of a species of bison supposedly extinct since the closing period of the Ice Age, has since been duplicated in other regions by scientists whose investigations were conducted under conditions imposing the strictest control. But the question of the exact age of the Folsom culture is still a controversial subject *revolving largely around the question of when these large Ice Age animals actually disappeared.* Most of the geologists and students of fossil animal life who have attempted to aid in dating the remains take the view that the last of these animals perished in the closing period of the last glacial retreat some 15,000 to 25,000 years ago. Some, however, express the belief that many of these archaic beasts lingered down into a period perhaps as late as 5000 B.C."

DECEMBER 1892

"'The intellectual honesty which is bred in the laboratory of chemistry and physics stands in strong contrast with the dangerous tendencies to plausibility, sophistry, casuistry, and self-delusion which so insidiously beset the pursuit of metaphysics, dialectics, and rhetoric. Much of the training given in college in my boyhood was, it is not too much to say, directed straight upon the arts which go to make the worse appear the better reason. Even where the element of purposed and boasted self-stultification was absent, there was a mischievous exaltation of the form above the substance of a student's work, which made it better to be brilliant than to be sound. Contrast with this the moral influence of the studies I am considering. The student of chemistry or physics would scarcely know how to defend a thesis which he did not himself believe. The only success he has hoped for has been to be right. The only failure he has had to fear was to be wrong. To be brilliant in error only heightened the failure, making it the more conspicuous and ludicrous.'—*Francis A. Walker, President of the Massachusetts Institute of Technology.*"

"By a very interesting, ingenious method and device invented by Mr. F. E. Ives, of Philadelphia, photographic pictures are secured by means of photographs taken on orthochromatic plates through selective color screens (*left*). From the triple negative thus obtained a positive transparency is made by contact, each picture and its several portions having the true color values. According to the modern theory of color vision, red, green and violet are considered the primary colors; consequently, the three pictures represent these three colors, and when viewed through an instrument provided with red, green and violet colored screens, and furnished with means for blending the three images into one, all the colors of the subject are shown."



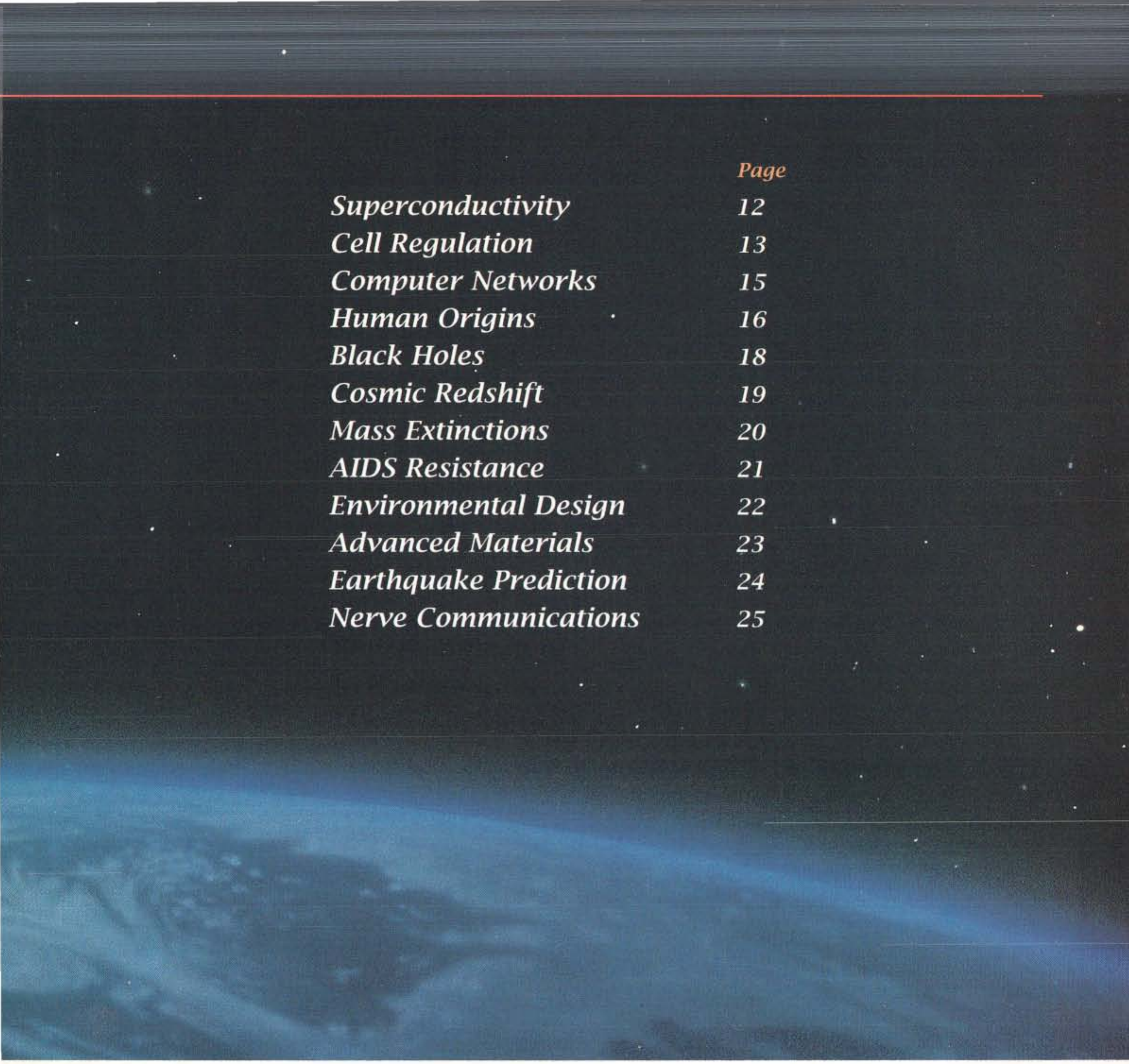
Color imaging in the heliochromoscope

The New Challenges

Are science's days numbered? To someone surveying the panorama of modern science, from astrophysics to zoology, the question must seem foolish. As the following reports demonstrate, scientists face countless challenges in pure and applied research. Some of these challenges were triggered by the discovery of phenomena—high-temperature superconductors, new neurotransmitters, an apparent periodicity in galactic redshifts—that upset established theories. Others stem from societal needs—AIDS vaccines, faster communications networks, environmentally benign products—that seem unlikely to be satisfied soon.

Yet talk of closure is in the air. With increasing conviction, scientists are asserting that they are converging on answers to some ultimate mysteries: How did the universe begin? How did life achieve its present diversity? What is the basis of consciousness? A slew of recent books by physicists—bearing such titles as *Dreams of a Final Theory* and *The Mind of God*—even suggest that physics might be approaching a fundamental description of matter and energy, sometimes called a "theory of everything."

At the same time, philosophers are insisting that, far from converging on the truth, science is degenerating into an ever more esoteric and fractious enterprise



ALFRED KAMAJIAN

that offers no coherent vision of reality. This trend is addressed in a book published this year called *The End of Science?*, which consists of papers presented at a conference at Gustavus Adolphus College in Minnesota in 1989. "There is an increasing feeling that science as a unified, universal, objective endeavor is over," the editors state. Some observers fear that this growing skepticism, coupled with the rising costs of science, could block its progress.

Neither of these scenarios offers much hope for the future of science—at least, science in its grandest sense, the quest to comprehend nature. Either the quest is ended by scientists' conviction that they have achieved final answers, or it is thwarted by political and economic forces. "Either way it's a tragedy," says Nicholas Rescher, a philosopher at the University of Pittsburgh.

Suggestions that science might be finite are commonly rebutted with references to 19th-century physicists, who allegedly believed Newtonian mechanics could explain and even predict all phenomena. The great British physicist Lord Kelvin is often said to have proclaimed that post-Victorian physics would consist of adding decimal points to the constants of nature—just before the theories of relativity and quantum mechanics shattered that illusion.

Could science be brought to an end by scientists' belief that they have final answers or by society's reluctance to pay the bills?

Actually, no evidence exists that Kelvin made such a statement, according to historians of science. Stephen G. Brush of the University of Maryland calls the complacency of 19th-century physicists "a myth" and says that most were aware of the limits of Newtonian physics. Lawrence Badash of the University of California at Santa Barbara doesn't go that far. Some Victorian-era physicists did in fact believe their profession had reached a kind of culmination. But this attitude was a "low-grade infection," Badash observes, rather than a full-blown disease.

Nevertheless, the epic progress of science in this century exerted a powerful influence on scientists' views of their enterprise. Many became convinced that the universe is both infinitely mysterious and infinitely knowable: the resolution of one question always generates still deeper questions. As the physicist John A. Wheeler of Princeton University has put it, "We live on an island of knowledge surrounded by a sea of ignorance. As our island of knowledge grows, so does the shore of our ignorance."

Others contend that the process of scientific discovery cannot continue forever. One of the most dramatic expressions of this view was made at the 1970 meeting of the American Association for the Advancement of Science. In a farewell address, Bentley Glass, the retiring president of the association, declared: "We are like the explorers of a great continent who have penetrated to its margins in most points of the compass and have mapped the major mountain chains and rivers. There are still innumerable details to fill in, but the endless horizons no longer exist."

Although few scientists would dare make such a sweeping pronouncement, many have implied that their particular fields might soon be reduced to the mere filling in of details. Particle physicists are most given to this kind of talk. In *Dreams of a Final Theory,* scheduled to be published this January, Steven Weinberg of the University of Texas predicts that within a generation or two physicists may be able to explain the origin and nature of mass and energy in terms of a single, unified theory. He even nominates a candidate: superstring theory, which traces all known forces and particles to infinitesimal loops of energy.

Weinberg, a Nobel laureate, takes pains to assert that the final theory of physics, whatever its form, will not bring all of science to a halt. Many higher-order phenomena, he notes, "from turbulence to thought," would remain to be investigated. But he argues that the theory would bring an end to "the ancient search for those principles that cannot be explained in terms of deeper principles."

The basic "rules" governing the universe may already be in hand, according to Frank Wilczek of the Institute for Advanced Study in Princeton, N.J. Although physicists will almost certainly continue to discover fascinating phenomena, such as high-temperature superconductors or buckyballs, that demand higher-order explanations, these findings are unlikely to change physicists' fundamental understanding of matter. Wilczek even downplays the importance of a unified theory. The standard model of particle physics, built on the firm foundation of quantum mechanics, "already describes matter under ordinary circumstances," Wilczek points out. "We're really talking about a few loose ends."

One loose end that may already have been tied up is the origin of the universe. Twelve years ago Alan H. Guth of the Massachusetts Institute of Technology drew on particle physics in proposing a creation scenario called inflation. This concept, which holds that the universe underwent an exponential expansion in its first microsecond, provides a remarkably specific model of the origin, present structure and ultimate fate of the observable cosmos.

Earlier this year investigators announced that the *Cosmic Background Explorer* (*COBE*) satellite had found evidence for inflation in a faint bath of microwaves thought to be the afterglow of the big bang. The physicist Stephen W. Hawking of the University of Cambridge proclaimed the finding to be "the discovery of the century, if not of all time."

Other workers harbor doubts about inflation, and cosmic mysteries still abound, such as the nature of the dark matter that, according to inflation (and other theories), dominates the universe. But theorists such as David N. Schramm of the University of Chicago expect these mysteries to be resolved within the context of the big bang theory. "Just because you can't predict tornadoes," Schramm has said, "doesn't mean the earth isn't round."

Biologists generally repudiate the possibility of complete descriptions of nature. "Biologists are too smart" to indulge in such predictions, says Ernst Mayr, an evolutionary biologist at Harvard University. Yet they have their own visions of completion. Mayr himself has indicated that a theory he helped to formulate—the so-called new synthesis, which consists of Darwinian evolution plus DNA-mediated genetic transmission—provides the basic framework needed to address all the major issues in evolutionary biology. All that remains, Mayr contends in *One Long Argument,* a book published last year, are "puzzles." Some of these puzzles, he concedes, particularly historical ones such as the origin of life or of *Homo sapiens,* are extremely difficult and may even resist a final, satisfying explanation. But however they are resolved, Mayr argues, they will not force any significant changes in the underlying paradigm of Darwinian evolution.

Most scientists would probably agree that the biggest challenges in biology—and in science as a whole—are those involving the brain and its faculties: perception, memory, emotion and ultimately consciousness itself. Only recently have these come to be viewed as legitimate topics of investigation. But researchers such as Francis Crick of the Salk Institute for Biological Studies in San Diego, co-discoverer of DNA's structure, are confident that the physical mechanisms giving rise to these ephemeral neural phenomena will eventually be unraveled.

Paradoxically, the rapid progression of science in this century—which fomented all this scientific hubris—led many philosophers to question the whole notion of scientific truth. Since the 1930s, Karl R. Popper of the London School of Economics has promulgated the view that theories are only provisional: they can never be proved but only disproved. Elaborating on this idea, Jeffrey Bub, a philosopher at the University of Maryland, suggests that scientists can approach the truth only by elimination; they can never say what the world is but only what it is not. Bub thinks this process—determining that "the world is not like that, it's not like that, it's not like that"—is inherently open-ended.

Even if the rules of nature are finite, like those of chess, might not science still prove to be an infinitely rich, rewarding game?

In 1962 Thomas S. Kuhn of M.I.T. went even further in challenging science's foundations. In his influential book *The Structure of Scientific Revolutions,* he argued that science is a highly subjective, political process rather than a rational search for truth. Lately Kuhn has attacked the notion of unification in science by asserting that it is dissolving into ever smaller fields whose worldviews are "incommensurable," or lacking a common unit of discourse.

Although Kuhn claims to support science, his writings have inspired a thriving antiscience movement, in which epistemological critiques are often coupled with moral and political ones. Even the old complaint that science is a blight on the human spirit appears to be gaining ground. This past year the British journalist Bryan Appleyard set forth this argument in a book entitled *Understanding the Present: Science and the Soul of Modern Man.* Although excoriated by scientists, the book has won praise from religious and even political leaders.

Such antiscience sentiments would not matter much if modern science were not so dependent on society's largesse. Paul Feyerabend, a philosopher at the Federal Institute of Technology in Zürich, thinks the most serious challenge to pure science is its rising cost. "This is a much greater danger than a few guys quacking about the legitimacy of science," he says. Society may grow increasingly reluctant to pay for expensive, pure-science projects such as the Superconducting Super Collider—and rightly so, Feyerabend says. "Is it more important to do this so people have more books [about physics] or to try and provide more food and medicine?" he asks.

Rescher, the University of Pittsburgh philosopher, notes that economic constraints are sure to worsen—at least for physics projects. "We can only investigate nature by interacting with it," he observes. "To do that, we must push into regions never investigated before, regions of higher density, lower temperature or higher energy. In all these cases, we are pushing fundamental limits, and that requires ever more elaborate and expensive apparatus."

Few philosophers think these problems will bring an end to science, but Gunther S. Stent, a biologist at the University of California at Berkeley, does. In his ironically titled book *The Coming of the Golden Age: A View of the End of Progress,* published in 1969, Stent contended that science and its by-product, technology, are doomed by their very success, their exponential rate of progress. Inevitably, Stent argued, given the limits of human resources and even cognitive capacity, our efforts will yield diminishing returns.

In addition, our achievements will cause our "will to power," our very desire further to understand and control nature—and even to engage in endeavors such as art and music—to wane. Instead we will turn to more hedonistic pursuits, perhaps even abandoning the "real world" for fantasies induced by drugs or electronic devices feeding directly into the brain. Virtual reality comes to mind. "Yes, I still think it will happen," Stent says now. Those who believe that science is a "limitless frontier," he adds, are "naive."

Some scientists express hope that ultimate theories will protect us from this fate by infusing the cosmos with meaning and purpose. Hawking, for example, has envisioned a unified theory of physics as a kind of mystical revelation that will allow humanity to "know the mind of God." Others, notably Weinberg, reject such sentiments as wishful thinking. Some 15 years ago Weinberg won some notoriety by asserting that "the more the universe seems comprehensible, the more it seems pointless." He notes that the statement has "dogged" him ever since, but he does not back away from it. "It would be wonderful to find in the laws of nature a plan by a concerned creator in which humans played some special role," he says. "I find sadness in doubting that we will."

Indeed, many scientists who believe in complete theories seem to view them with ambivalence. David Gross of Princeton, a superstring theorist, has dedicated his life to finding a theory that will answer "all the questions that can be asked" by physicists. Yet part of him dreads the possibility that he and his colleagues might actually reach that goal. "Nobody worries about completion in the sense of losing their jobs," he remarks. "But I think there would be a feeling of loss."

Complete theories of human cognition and behavior might offer even less solace, according to the biologist Edward O. Wilson of Harvard. "When we have progressed enough to explain ourselves in...mechanistic terms," Wilson wrote in *Sociobiology,* published in 1975, "the result might be hard to accept." Wilson ends his book with a quote from the French existentialist Albert Camus: "...in a universe divested of illusions and lights, man feels an alien, a stranger. His exile is without remedy since he is deprived of the memory of a lost home or the hope of a promised land."

Fortunately, science will be complete only when scientists believe that it is, and their research offers potent counterarguments to that possibility. For example, recent studies of so-called chaotic and complex phenomena—from gushing faucets to stock markets—strike a blow against the facile reductionism underlying many predictions of completion. "As one ascends in levels of complexity from quarks to human societies, one finds properties that cannot be predicted from the properties of the parts," explains Stuart A. Kauffman, a biochemist at the Santa Fe Institute, a center for complexity work. Such studies, he adds, show that there is "no finite way of parsing the world into objects and laws by which they interact."

Mathematics also undermines the concept of completeness, according to the physicist Freeman J. Dyson of the Institute for Advanced Study. In the 1930s, he notes, the mathematician Kurt Gödel demonstrated that no finite set of axioms can answer all the questions that it raises; mathematics, in other words, is infinite. Dyson thinks the same is almost certainly true of physics. "It's foolish to imagine anything in the nature of a culmination," he says. "That's a very parochial view."

Another possibility exists. Perhaps the rules governing nature—the knowable rules—are finite in number. Even so, science may remain an infinitely rich, rewarding enterprise. Making this point, Wilczek, Dyson's colleague at the Institute for Advanced Study, likens us to pawns on a chessboard. Our first task is to figure out the rules of the game. Once we know the rules, we can transform ourselves from pawns to players.

In a sense, this process has already begun: artificial intelligence, genetic engineering and other technologies may represent our first tentative steps toward becoming players. And if chess permits a virtually infinite variety of games, the rules of nature surely do. Science may be immortal after all. *—John Horgan*

Theoretical Resistance

What makes some ceramics superconductive?

Six years ago this month physicists received a holiday gift that few probably had on their wish lists. Researchers in Japan confirmed what J. Georg Bednorz and K. Alex Müller of the IBM Zürich Research Laboratory had found: bits of potterylike material can conduct electricity without resistance at the reasonably balmy temperature—at least to physicists—of about 30 kelvins.

Yet despite all the intense research that followed, no one knows what causes the phenomenon. Unlike the materials themselves, every theory of high-temperature superconductivity seems to have some resistance. "They are very strange materials that present us with very strange data," says Patrick A. Lee, a physicist at the Massachusetts Institute of Technology.

The superconducting ceramics consist of parallel layers of copper-oxygen mixed with such elements as yttrium, barium or lanthanum. There are about 40 such copper oxide compounds, or cuprates, each with different superconducting transition temperatures. The present record for the highest critical temperature belongs to a thallium-based cuprate that begins to superconduct at 130 kelvins (-143 degrees Celsius).

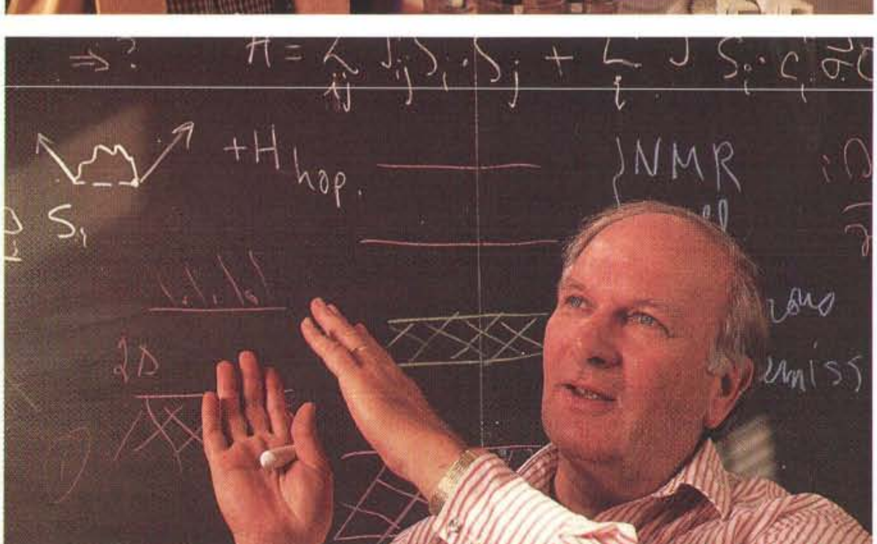
Exactly why the electrons should zip unencumbered between the copper-oxygen layers remains a mystery. The workhorse of superconductivity theory has been the BCS model, proposed in 1957 by John Bardeen, Leon N. Cooper and Robert Schrieffer to explain the behavior of classic, low-temperature metallic superconductors. In this scenario, electrons are scattered by phonons, which are vibrations in the lattice of the material. Although electrons have the same charge and would normally repel one another, the scattering by phonons induces the electrons to "pair bond." Bonded together, these so-called Cooper pairs of electrons can slip through the conductor without resistance.

The cuprates, however, were a surprise. Although the electrons do pair up in the copper oxides, "the BCS model could not possibly apply," says Princeton University theorist and Nobel laureate Philip W. Anderson. At the observed transition temperature, the lattice vibrations should have disrupted the supercurrent flow. "People have tried to pull and tug and squeeze the BCS model," Anderson remarks. "All they've done is confuse themselves."

Part of the difficulty in formulating a successful theory is that it is tough to tell exactly what cuprates are. One approach is to look at the cuprates as a "normal" metal. Even though they are brittle and stiff, cuprates do, after all, conduct electricity. According to seminal work by the Russian physicist Lev D. Landau, the behavior of ordinary metals is described by Fermi liquid theory. This formulation explains the low-temperature behavior of metals by ignoring the charge repulsion between electrons. Landau argued that electrons can be replaced by "quasiparticles," which do not strongly interact with one another. (To experimenters, these quasiparticles are just electrons.)

Unfortunately, the properties of normal metals are not evident in experiments conducted with cuprates. One


ROBERT PROCHNOW



PHIL SEARS

CURRENT THINKING: Philip W. Anderson of Princeton University (top) *wants to revamp conventional models to explain high-temperature superconductivity. Robert Schrieffer of Florida State University* (bottom) *believes only a few key modifications are needed.*

type of experiment, for example, looks at the Hall effect—the sideward drift of electrons as they flow through a metal in a magnetic field. The drift creates an electric field called the Hall field. In metals the Hall field is independent of temperatures above a threshold of about 100 kelvins. In contrast, the Hall field in cuprates shows a pronounced temperature dependence above the transition point. Other experiments conducted this year support the view that the cuprates are much more than conventional metal theory can describe.

Theorists have wended their way among several ideas to account for the experimental results and to formulate a model that would explain how the needed electron pairing happens. Schrieffer, now at Florida State University, thinks the answer "is very close to the usual pairing theory of superconductivity." In general, he says, the cuprates are ordinary metals possessed of a magnetic property that forces the spins in electrons in neighboring copper atoms to align oppositely. Phonons play only a limited role in setting the critical temperature in cuprates. Schrieffer believes quasiparticles pair up in a "spin bag," a theoretical construct he likens to a rubber sheet. A ball placed on this sheet will deform it. A second ball will be attracted to the indentation and roll into it, forming a bonded pair.

While Schrieffer tries to add chapters to superconductivity theory, Anderson wants to rewrite the book. In his view, the normal state of the materials is radically different. Anderson breaks up two properties of quasiparticles—charge and spin—into different particles: spinons and holons. Spinons would be neutral particles that have spin. Holons would have a positive charge but no spin. This theoretical holon-spinon liquid has unusual properties, one of which is that the liquid is strictly confined to between the copper-oxygen planes. Above the critical temperature, the confinement would greatly hinder any conduction perpendicular to the planes (the c axis)—one of the well-known and still poorly understood properties of the cuprates.

In his model, Anderson says, "the interaction responsible for the high-transition temperature is essentially a recovery of the conductivity in the c axis." Calculations show that only pairs of electrons can tunnel between planes; single electrons cannot. In the superconducting state the cuprate becomes a three-dimensional metal through which paired electrons travel. Conveniently, "my mechanism of coupling between layers acts in parallel with any BCS effects," Anderson maintains. The theory is "so blatantly obvious that it hits you in the face."

Well, not everyone's face. Several investigators have modified some of Anderson's insights because the calculations are completely solvable only in one dimension. "That two- and three-dimensional systems should behave like one-dimensional ones is unproven to me," says Chandra M. Varma, an AT&T Bell Laboratories theorist. Varma tempers Anderson's basic ideas, representing the cuprates as a "marginal Fermi liquid."

Other researchers, including Lee of M.I.T., have applied quantum mechanical calculations to Anderson's fundamental concepts. The resulting formulation is called the t-J model (t refers to the kinetic energy of the electron, and J is the interaction between the spins of the electrons). According to this model, the question is really not why the transition temperature is so high, but why it is so low. The t-J model indicates that superconductivity is suppressed by fluctuations, or variations in the movement of the electrons.

The persistence of so many theories may say as much about personalities and the absence of critical data as it does about the inherent nature of the cuprates. "Everyone has staked a claim," observes N. Phuan Ong of Princeton University. And everyone seems able to find evidence that supports their views and discredits those of the others. Anderson's model may not go far enough. Electrons in spin bags may actually repel one another. Varma's ideas do not precisely detail the coupling phenomenon. Calculations from the t-J model require several approximations to yield the right answers.

Ong leans toward Anderson: "A lot of experiments seem to confirm what he says." One such piece of evidence comes from Osaka University, where a calcium- and strontium-based copper oxide seems to have a critical temperature near 160 kelvins, the highest to date. If confirmed, the finding would accord with Anderson's calculations.

This compound suggests that superconductivity theories could be predictive—a large step from the standard BCS model, which was only descriptive. "Suppose the mechanism is as Anderson said," Ong says. "Then you would search for a compound that has the largest number of adjacent copper-oxygen layers and the tightest coupling between layers." In other words, the highest critical temperatures would presumably occur when all the layers are identical and equally spaced. Indeed, the putative 160-kelvin superconductor is just such an "infinite layer" compound. (More familiar cuprates have only two or three adjacent layers.)

A triumphant theory would indicate a direction for developing new superconductors, and experiments during the next year should narrow the theoretical field. Results have already buried some previously popular ideas, such as the notion that superconductivity is induced by particles called anyons. But whichever theory ultimately prevails, the physicists agree on one point. The cuprates are unlike anything they have seen before—not an insulator but not quite a metal. "My own feeling is that the materials are changing the future view of the domain of condensed-matter physics," Schrieffer says. "Textbooks need to be rewritten." *—Philip Yam*

Essential but Expendable

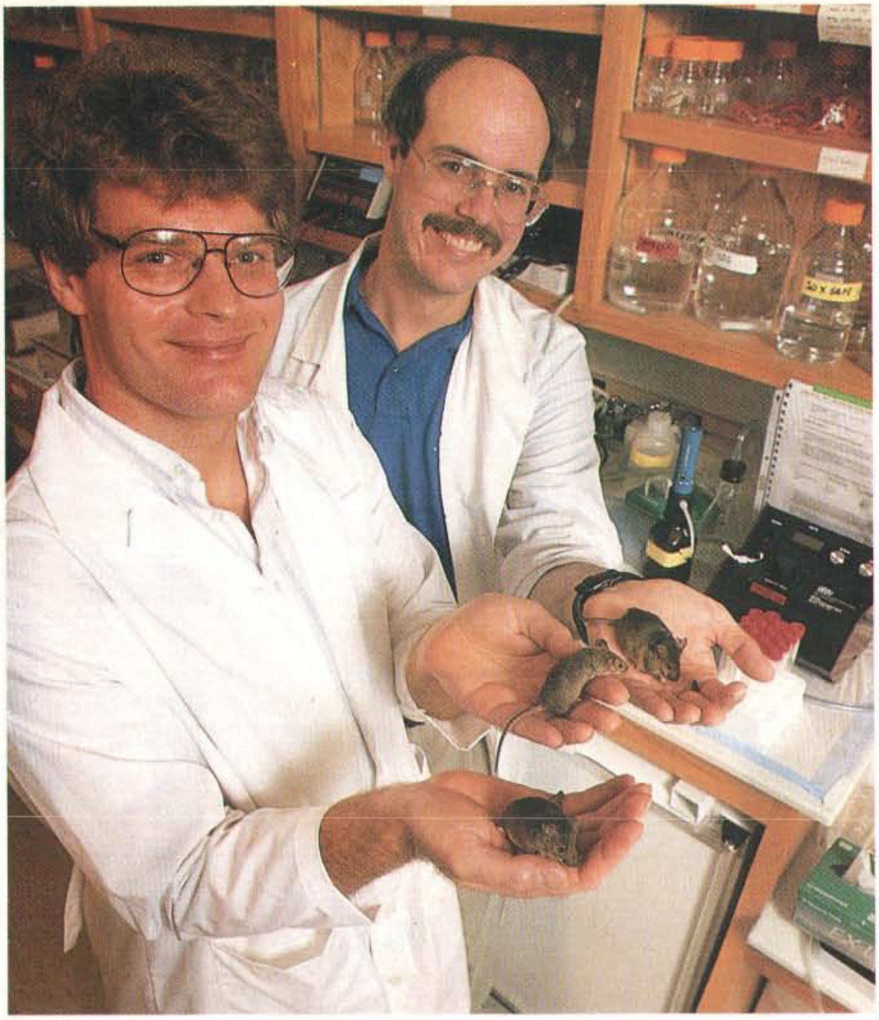
When do master genes regulate cell growth?

Geneticists buzz about knockouts these days more often than most boxing fans do. Using molecular probes, experimenters can obliterate individual genes in mice to see how the loss hurts the organism. This year Lawrence A. Donehower and Allan Bradley of the Baylor College of Medicine and their colleagues took on *p53*, a gene that looked like a heavyweight in embryonic development. To the surprise of most spectators, even after the gene was down for the count, the mice were still standing. "I think it caused a lot of people to reassess the role of *p53*," Donehower says.

That finding and a host of other recent experiments, including some concerning a gene designated *Rb*, are providing new clues about what regulates cell growth. Some biologists are now reexamining hypotheses about the irreplaceability of the genes during development and their role in cancer.

Excitement about the observation stems from the fact that both *p53* and *Rb* are tumor suppressor genes. When they are absent or damaged in animals and people, tumors tend to develop. Individuals who inherit only one of the usual two *p53* genes manifest Li-Fraumeni syndrome, a predisposition to acquire cancer of the breast and many other tissues. People with only one copy of *Rb* often develop retinoblastoma, a tumor of the retina for which *Rb* is named.

Such evidence led many cell biolo-


DAN FORD CONNOLLY

KNOCKOUT MOUSE is held by its creators, Lawrence A. Donehower and Allan Bradley of the Baylor College of Medicine.

gists to assume that *p53* and *Rb* were master genes that choreographed the growth and division of cells during the orderly formation of an embryo. Without them, the reasoning went, an embryo might turn into a disorganized mass, never to be born. When Donehower and Bradley used knockout technology to deactivate *p53* in mice, however, the animals survived and looked completely normal at birth. Yet more than 70 percent of them developed tumors within six months. In short, *p53* seemed to be dispensable during early development but essential for preventing tumors later.

One scientist who was not surprised was Michael B. Kastan of the Johns Hopkins University School of Medicine. In August he presented evidence that *p53* acts as a "checkpoint," or monitor, of DNA integrity. Kastan found that after a dose of gamma rays, normal cells arrest their growth, apparently to give repair mechanisms time to undo the genetic damage done by the radiation. The p53 protein acts as "a signal to stop synthesizing DNA until the damage is repaired," Kastan says. "If *p53* is abnormal, that signal isn't there, and the cell continues to replicate a damaged DNA template." As the damage mounts, some cells undergo changes that transform them into tumor cells.

Experiments from the laboratories of Geoffrey M. Wahl of the Salk Institute for Biological Studies in San Diego and Thea D. Tlsty of the University of North Carolina at Chapel Hill have verified that *p53* prevents at least one type of genetic abnormality common in tumors. Wahl disagrees with Kastan's interpretation of *p53*'s role, however; he argues that *p53* is not responding to genetic damage so much as it is to shortages of certain metabolites essential to gene function and replication.

Bert Vogelstein of Johns Hopkins, a pioneer in the study of *p53*, emphasizes that neither the knockout mice nor the other work settles the fundamental problem of how *p53* acts in embryos and tumors. For example, the knockout mice might have developed correctly because other genes took over for the missing *p53*. "Many genes that one might have thought would be critical for normal development can be knocked out, and the mice do fine," Vogelstein says. "Mammals have an incredible capacity for tolerating losses of specific genes."

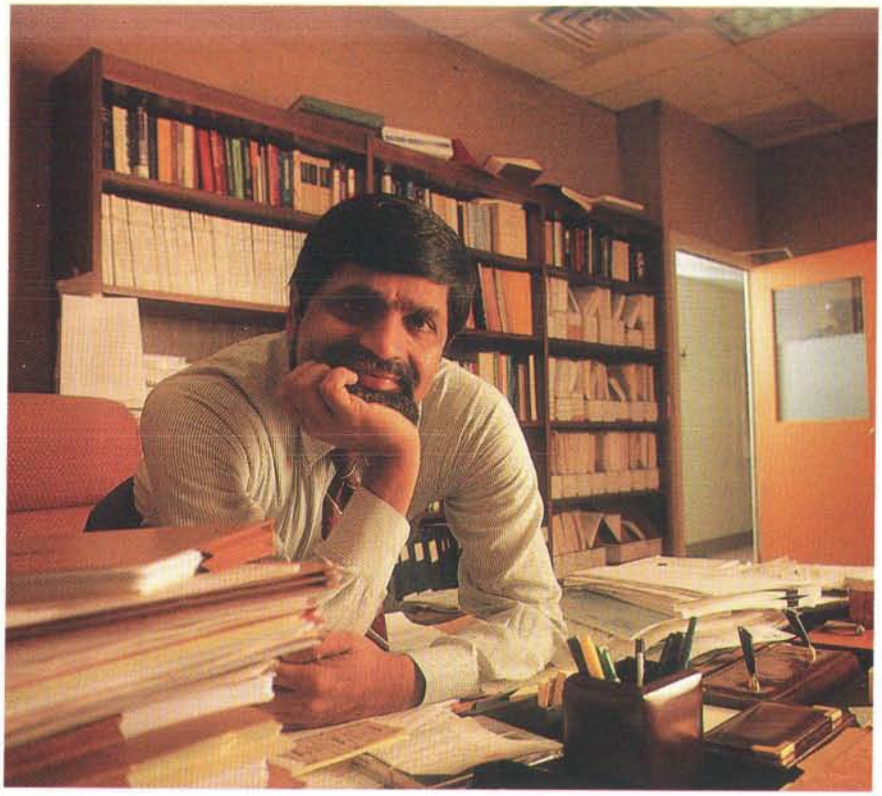
Bradley doubts that *p53* is genetically redundant. Rather it may have only a highly specific function, and its absence may produce only highly specific defects. Indeed, he thinks that in many cells containing the p53 protein, it may have no real function at all. "We have junk DNA. Why can't we have junk RNA and junk proteins in a cell?" he asks.

For whatever reason, the elimination of *p53* does not destroy mouse embryos. Knockouts of *Rb* tell a different story. In separate studies, Bradley, Robert A. Weinberg of the Massachusetts Institute of Technology's Whitehead Institute and Hein te Riele of the University of Amsterdam found that mice without *Rb* never grow past the 16th day of embryonic development. Instead their nervous systems and blood cells fail to differentiate properly. Mice that have only one *Rb* gene develop normally but often die of brain tumors at an early age.

Yet even those results raise questions about the role of *Rb*. The lack of *Rb* does not trouble the embryos until about 13 days into their development. What was regulating the cell cycle before then? Also, as Bradley points out, *Rb* is expressed in virtually all cells of the body. Why, then, do only a few tissues develop abnormally when the gene is not present?

These are only a few of the questions steering further research on *Rb* and *p53*. Far more work is needed to determine precisely when, why and how both genes become active. Much of it, Vogelstein says, will concentrate on finding the genes and proteins controlled by *p53* and *Rb*.

The work on *p53* could have a direct bearing on the treatment or prevention of cancer. Kastan suggests that in future therapies, cancer patients might receive a low preliminary dose of radiation to stop the growth of normal tissues in the body temporarily. Because the cancer cells without *p53* would grow heedlessly, chemotherapy drugs that kill growing cells would affect them more selectively. Donehower also notes that because of their sensitivity, mice without *p53* may be useful for testing the carcinogenicity of substances. It appears that, notwithstanding their knockouts, these mice could still be champions. —*John Rennie*



ROBERT PROCHNOW

ARUN N. NETRAVALI of AT&T Bell Laboratories sees a new paradigm of communications emerging from the fusion of computers and telecommunications.

Racing Light

Can computer networks handle the traffic?

For the past few decades, such visionaries as Aldous Huxley, Marshall McLuhan and Vannevar Bush have been preaching about a time when any kind of information—data, text, pictures, audio and video—would be as readily available as tap water. To that end, technology pioneers knew they needed robust communications links. But for the most part, they believed the existing telephone and computer networks offered a sturdy infrastructure for the future superhighways for data.

Now they are facing gigabit gridlock. "We're on the threshold of a technical discontinuity," says Richard D. Gitlin, who heads the network systems research department at AT&T Bell Laboratories. Arun N. Netravali, executive director of Bell Labs' communications sciences research division, cites the recent spate of network failures as a symptom of the problem. "We need a new paradigm for computer network communications," he asserts.

To fend off impending disaster, telephone companies are confronting the fact that the schemes they have traditionally used to switch voice communications are no longer adequate. The computer community realizes demands have outgrown the hardware and software strategies it uses to weave networks over small areas. And together computer and communications experts are thinking through what happens when almost any aspect of a network can change—except the speed of light.

The idea that computer and communications researchers would join forces to design very high speed networks might sound logical, but it flies in the face of more than two decades of rivalry. Telephone companies still handle almost all calls using a strategy called circuit switching, the electronic analogue of the tasks performed by early switchboard operators who physically plugged one caller's line into another. A voice conversation takes up a predictable amount of room, or bandwidth, on a network and involves immediate exchanges of information. Because circuit switching offered dedicated channels for transmitting continuous, real-time streams of information, it was well suited for sending voices across the ether.

When computers began talking to one another in the 1970s, however, they made very different demands on the communications links. To those who wanted to move megabits of data, the 64-kilobit-per-second bandwidth of the most advanced voice networks looked like a rutted country road. Rather than rely on the fixed bandwidth connections set up by circuit switching, computer engineers devised techniques for sharing bandwidth by sending information in irregularly sized bursts and letting multiple machines jointly make use of the available channels by interleaving bursts. Transmitted data were divided into chunks, or packets, each of which included details about the contents and destination of a message. To minimize the time such packets spent in transit, the engineers put a heavy burden of processing on the switches responsible for routing the data.

Nevertheless, the differences between the two camps began to fade when computer and communications engineers realized they both wanted to send the same kind of information—from data to video. "As you go to high speeds, packet switching becomes more appropriate for a wide spectrum of services over the nets," says Inder S. Gopal, who manages broadband networks at IBM.

Gopal believes telecommunications may soon wind up looking like distributed computing. Although 99 percent of the telephone networks are still based on circuit switching, telephone companies have pledged to adopt packet switching. "As we scale up to 150-megabit-per-second networks, we will switch to packets," says W. David Sincoskie, executive director for computer networking research at Bell Communications Research (Bellcore), the research arm of the regional telephone companies.

In the mid-1980s the communications industry began codifying so-called Asynchronous Transfer Mode (ATM) protocols, in hopes of creating an international standard for high-speed packet switching. Yet traditional packet switching has its faults. Because users will expect the same kind of instantaneous response they receive in voice calls, packet-switched networks must simulate the feel of a dedicated connection.

As a result, engineers have begun to grapple with the bottlenecks caused by the heavy processing burden they have loaded onto their switches. "All the research of the past was aimed at optimizing the bandwidth," Gopal explains. Fiber-optic networks, capable of shipping tens of terabits of information per second, eliminated that constraint. Now every extra nanosecond a switch spends investigating a packet is time lost.

"Even with a high-speed network, you can't change the speed of light," Netravali notes. When signals are sent from, say, New York to California along gigabit links, half a million packets may be

in transit before the first packet is received, much less checked.

To circumvent processing jams, Gopal says, engineers are designing "lite" protocols. For instance, messages may be checked only when they reach their final destination, not along the way. And rather than rely on fast, albeit general-purpose, computers as switches, companies are building specialized devices. Among the contending switch designs are "Planet," an IBM effort, and "Sunshine," developed by Bellcore.

After several years of exploring these issues in laboratories, researchers are getting a chance to put their switches and protocols through their paces. In conjunction with industry, the government is funding a series of experiments aimed at laying the groundwork for a multiple gigabit-per-second National Research and Education Network. A year of tests is slated to begin shortly. Although companies, including IBM and AT&T, are running additional high-speed networking trials, the gigabit test-bed—coordinated by a tiny, not-for-profit company, the Corporation for National Research Initiatives (CNRI) in Reston, Va.—will be the most visible forum.

The fusion of computing and communicating technologies taking shape in the gigabit test-bed promises to unleash a sprawling set of challenges. For instance, how can computers take advantage of the gigabits of data churning through networks, asks Robert E. Kahn, founder of CNRI. Netravali is looking for new multimedia-based architectures that will not have to examine the incoming flood bit by bit. Finding enough addresses to connect all interested parties to the network is also no small feat.

Such questions will continue as researchers free up yet more bandwidth. Beyond speeds of five gigabits per second, Gopal says, designers will have to find ways of pushing the electronic switches to the very edges of networks—or eventually eliminate them altogether. "All-optical" networks, for instance, might rely on different wavelengths of light to transmit information. At Bellcore and AT&T, researchers have devised arrays of minuscule lasers that emit a spectrum of wavelengths. An IBM team has been working with receivers that can be rapidly tuned to pick up only particular frequencies.

Technology permitting, investigators will continue to reach for the goals envisioned by the prophets of the age of information technology. Kahn is optimistic that the scientific community will make fruitful use of the new technologies. "After all," he quips, "'How you gonna keep 'em down on the farm after they've seen Paree?'" —*Elizabeth Corcoran*

Shaking the Tree

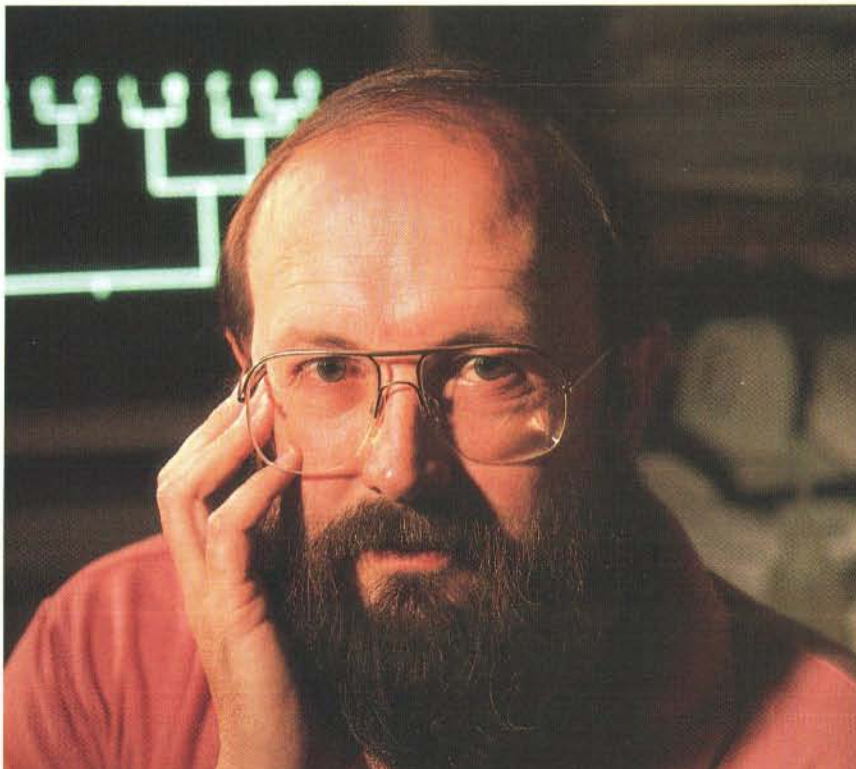
Will statistical analysis of DNA pinpoint human origins?

We know that the genus to which we belong originated in Africa. Did our species emerge there, too? Five years ago a group of geneticists under the late Allan C. Wilson of the University of California at Berkeley thought it had established this second African root by constructing a family tree that traced modern *Homo sapiens* to a woman who lived in Africa some 200,000 years ago. Then, less than a year ago, statisticians claimed they had felled that tree when they demonstrated a plethora of equally plausible origins in Asia. Now the proponents of African Eve, as she is called, are refining their methods and seeking new lines of evidence.

The boldness of the Wilson group's approach consisted of exploiting the one relevant fossil that was present in magnificent abundance: the DNA in the cells of living people. The researchers compared mitochondrial DNA, which passes from mother to child as a single genetic unit. Because every variation in each independently inherited unit of that DNA traces to a single prototype, one can plot it as a tree that would correlate with the tree of human populations. The populations on the deepest branches would point to the root of the tree—the homeland of humankind. The workers dated the root by comparing the differences among human DNA types with those that have accumulated between humans and chimpanzees in the five to eight million years since they diverged.

Critics of the model consist of two categories: those who made themselves heard right away and those who waited. The audible group was led by Milford H. Wolpoff of the University of Michigan, a paleontologist who favors the alternative model, which posits a multiregional emergence of *H. sapiens*. According to this view, modern humans evolved from earlier populations in many regions at once, over hundreds of thousands of years. Wolpoff and other proponents profess to see signs of racial continuity between archaic and modern populations. They thus deny the Africanists' contention that a small band of interlopers replaced earlier populations and only then differentiated to form modern racial groupings.

The other category of critics treaded carefully, in part because they came from outside anthropology. The major challenger, Alan R. Templeton of Wash-


DAVIS FREEMAN

THEORETICAL GENETICIST ***Joseph Felsenstein*** *of the University of Washington judges an evolutionary model by sampling the thousands of family trees it allows.*

ington University, a theorist of phylogenetic reconstruction, has still not published his full-scale argument. Templeton's most stinging blow, delivered in preliminary form in *Science,* concerns the construction of Eve's genealogy. He has shown—as have other geneticists working independently—that Wilson's group misapplied the computer program they used to create the tree. With the same software, Templeton found thousands of trees at least as short as Eve's, many of which had their longest branches in Asia instead of Africa.

Templeton said he became suspicious of the Berkeley work when he realized how heavily it depended on a single genetic unit. "But it turns out that the details of the population tree have nothing to do with the out-of-Africa theory," he says. Templeton argues that there are no statistical traces of the demographic expansion that should have followed the movement out of Africa, just small, local expansions in Asia and Europe. "It disproves the African replacement hypothesis—in fact, any replacement hypothesis," he asserts.

Among those most disappointed by Templeton's analysis was Christopher B. Stringer of the Natural History Museum in London. As an original framer of the out-of-Africa theory and one who depended on paleontological evidence alone, he admits he was probably seduced by the geneticists' support. "Too much certainty was claimed for the original genetic data—a number of us were misled, and I was one of them," Stringer acknowledges. Yet he insists that the preponderance of evidence, genetic and paleontological, still points to Africa.

For their part, Eve proponents admit their computer error but refuse to throw in the towel. Mark Stoneking, a Berkeley veteran now at Pennsylvania State University, stands by the replacement proposition. "Templeton's using the wrong data base," he argues, and that is why he finds no statistical signal. "His base has more populations in it but is based on low-resolution restriction maps. We are using sequencing data—that's the highest possible resolution—and find that every human population shows signs of expansion, occurring at about the same time and in the same high degree."

Going on the attack, Stoneking and Linda Vigilant, also formerly of Berkeley, have just published in the *Philosophical Transactions of the Royal Society* evidence contradicting the multiregional hypothesis. They link their DNA clock not to the split with chimpanzees but to a more accurately datable event: the human settlement of New Guinea some 40,000 to 60,000 years ago. The differentiation since then gives a rate of change that could place Eve between a lower boundary of 63,000 years and an upper boundary of 416,000 years ago. The range has 95 percent reliability, the standard statistical test of such measures. "Therefore, non-African populations that date back more than 416,000 years could not have contributed any mitochondrial DNA types to modern populations," the authors conclude. The multiregional theory assumes, however, that today's races began differentiating about a million years ago.

But you can wring only so much information from the mitochondria. Other genetic loci are now coming under close study, and the resulting information will itself be subjected to more sophisticated tree-building techniques. Kenneth K. Kidd of Yale University has investigated the Y chromosome. "We are interested in such questions as whether there is greater genetic diversity in Africa than outside of it," he explains. Robert Dorit, also of Yale, says the chromosome's lack of variation, or polymorphism, "is compatible with a very recent origin but does not confirm it."

To hack their way through the forest of trees, geneticists are trying new techniques. Joseph Felsenstein of the University of Washington is refining maximum likelihood analysis. This method computes the total probability of the observed data to evaluate the assumptions of an evolutionary model. One might test the out-of-Africa model, for instance, by plugging in its predictions for the size of ancient populations, their rate of migration and so forth, then seeing whether the trees produce data sets resembling those of the modern world. "We think we can handle some samples on the order of the size seen so far," Felsenstein says. "We would perhaps be talking about 100,000 trees—a week on a modern workstation. Within the next year, we should be able to evaluate the actual data."

In the future, he adds, the emphasis will be not on analyzing one gene to exhaustion but on analyzing gene after gene. If modern populations stem from Africa, so should every one of their genes. No single brilliant stroke will cut the Gordian knot, as so many anthropologists hoped when Eve first seemed within their grasp. *—Philip E. Ross*

Puzzling Powerhouse

Black holes power the brightest galaxies, but how?

What sits at the center of a galaxy, devours millions of stars during its lifetime, spews out more radiation than a billion suns and generates jets of particles that extend for thousands of light-years? Any astronomer can tell you it is an active galactic nucleus (AGN)—essentially a massive black hole surrounded by a rotating cloud of gas. Yet some three decades after the discovery of AGNs, astronomers remain perplexed about just how black holes function as galactic engines.

Black holes are believed to be the densest objects in the universe, exerting gravitational forces so great that anything that gets too close is forever imprisoned. To be sure, there is no direct evidence that black holes even exist. But in recent years, astronomers have been building a strong case that black holes are the energy source for AGNs. "Its hard to imagine that AGNs are powered in any other way," comments John L. Tonry of the Massachusetts Institute of Technology.

Three lines of reasoning support the case for black holes. First, AGNs can vary enormously in brightness over a period of days. This observation indicates that the power source of the AGN can be no bigger than the distance it takes light to travel in a few days. The source must therefore be both energetic and compact. Second, astronomers can measure the speed with which stars orbit around the source. Such measurements reveal the strength of the source's gravitational pull and therefore its mass. Last, an AGN emits so much radiation over such a wide range of wavelengths that it cannot possibly consist of ordinary stars—not suns nor giant stars. The most likely candidate is a black hole surrounded by a cloud of gas because it would radiate everything from microwaves to gamma rays.

By applying these three arguments and others, astronomers have recently found "very convincing" evidence for dozens of black holes in active galactic nuclei. "Astronomers are 99.9 percent sure that black holes exist," Tonry says. For example, using the *Hubble Space Telescope,* Tod R. Lauer and his colleagues at Kitt Peak National Observatory found evidence for a black hole in the giant elliptical galaxy M87. And in July, John Kormendy of the University of Hawaii and Douglas Richstone of the University of Michigan announced the discovery of what must be one of the largest black holes. It is more than a billion times the mass of the sun.

To date, astronomers have figured out how the gravitational energy of black holes is converted into tremendous flows of radiation, but they are mystified by many of the details. As gas spirals down toward the hole, it accelerates almost to the speed of light. The matter becomes extraordinarily hot as a result of friction and electromagnetic forces, and it therefore radiates intensely. A black hole should transform matter into energy with more than 10 times the efficiency of ordinary stars. Yet how does a black hole in an AGN attract enough gas to satisfy its voracious appetite, and how does it manage to generate long jets of particles?

A black hole can consume billions of stars, or the equivalent amount of gas, during its lifetime. Presumably, the gravitational field of a black hole can capture stars or other matter speeding through a galaxy. The imprisoned objects then orbit around the hole at some distance. Indeed, they would do so indefinitely were it not for occasional friction. But astronomers have difficulty devising a mechanism that produces enough friction so that at least one star, or the equivalent, falls into the hole every year for a billion years. "There are no good answers," admits theorist Julian H. Krolik of Johns Hopkins University.

Workers have somewhat more success explaining the formation of jets. According to Roger D. Blandford and his colleagues at the California Institute of Technology, jet formation is intimately linked to the magnetic fields that are produced in the gas clouds rotating around the black hole. The magnetic fields get twisted up in the cloud, and this twisting generates a force that opposes the rotation of the gas. Consequently, the gas slows down and falls toward the black hole. As it does so, it is swept up in an electromagnetic wind and is ultimately shaped into jets.

To prove this theory, astronomers need to make a series of observations that are "all quite hard," Blandford says. For example, researchers might find clues by using radio telescopes to resolve the finest structures in AGNs. Theorists have recently benefited from the precious data streaming from the satellite known as the *Compton Gamma Ray Observatory*. The results indicate that the jets of AGNs are strong emitters of gamma rays and that particles in the jet travel at speeds approaching that of light. Such observations have ruled out some of the more conventional theories for jet formation while constraining others. "Most of us have the confidence," Blandford remarks, "that we will have much more definitive answers in five years' time." —*Russell Ruthen*

Quantum Dissidents

Is there unexpected order in the cosmos?

In 1929 Edwin P. Hubble discovered that the light from a distant galaxy is reddened in proportion to the distance of the galaxy from our own. He and most astronomers since have interpreted this shift as a cosmic Doppler effect: the frequency of the light is slowed for the same reason that the whistle of a train moving away from a listener is lowered in pitch. Hubble concluded that galaxies are moving away from telescopes, the more distant galaxies receding faster than nearby ones. The implication that the universe is expanding has become the dominant theory of cosmology—almost the only theory.

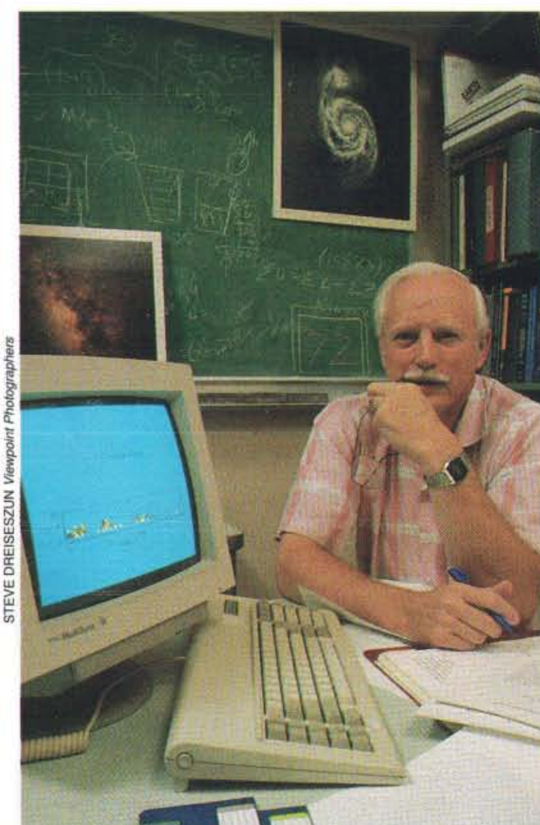
A few dissident astronomers do not accept that redshifts are simply the velocity of a receding galaxy. One is William G. Tifft of the University of Arizona. Statistical analysis has convinced Tifft and his colleague W. John Cocke that redshifts of different galaxies are not spread smoothly over a range of values, as under the standard interpretation. Instead they believe redshifts are "quantized"—they tend to fall on evenly spaced values, like the rungs of a ladder.

While many cosmologists rejected Tifft's hypothesis out of hand, some astronomers set out to prove him wrong. Their efforts have been disappointing. Cocke had expected to refute Tifft when he started studying redshift data more than a decade ago but instead ended up joining forces with him. And a new analysis by Bruce N. G. Guthrie and William M. Napier, who worked until recently at the Royal Observatory in Edinburgh, appears to give Tifft and Cocke more substantial support. "I thought I was going to debunk it," Guthrie says. "I was quite surprised to find the hypothesis held up."

If Tifft and Cocke's ideas are right, the implications for cosmology would be profound. It could mean, for example, that the main foundation of the subject—that redshifts of objects correspond to their recessional velocities—is wrong. Or it could mean that some unrecognized type of quantum effect operates at the level of very large objects as it does on very small ones, constraining redshifts. Guthrie speculates that redshift quantization may be a relic of interactions between light and matter in the early universe.

Tifft started to form his ideas in the early 1970s. He noticed that galaxies in one particular cluster fell into discrete groups when classified by redshift and by brightness. Subsequently, he analyzed radio telescope data gathered by others and found that redshifts measured from the two members of a galaxy pair, or from the various members of a galaxy cluster, tended to differ by multiples of 72 kilometers per second Even redshifts measured within a single galaxy seemed to show the effect.

Later studies on widely separated galaxies started revealing steps, or periodicities, of less than 72 km/sec, so Tifft extended the hypothesis to simple fractions of that number. He found a statistical periodicity of 24.15 km/sec in redshifts of dwarf galaxies and another one of 36.2 km/sec among "wide profile" galaxies. Those numbers are about one third and one half of the basic 72 km/sec figure.



STEVE DREISESZUN *Viewpoint Photographers*

WILLIAM G. TIFFT of the University of Arizona thinks redshifts evince a physical phenomenon overlooked by other astronomers.

Although an independent study in 1989 found weak support for a 36.3-km/sec quantization among wide-profile galaxies, most astronomers remained deeply skeptical. Then Guthrie and Napier, who had failed to confirm Tifft's hypothesis of a 24.15-km/sec redshift period among dwarf galaxies, examined data from 89 single spiral galaxies that had not previously been studied. That analysis was more convincing.

Single-galaxy redshifts are harder to calculate than differences between members of a cluster, because allowance has to be made for the motion of the earth and of the sun. (The earth's motion cancels out in comparisons between members of a cluster.)

What Guthrie and Napier found was a seemingly strong periodicity of 37.2 km/ sec—close to Tifft's number of 36.2 km/sec—that appeared only when appropriate corrections were made for the motion of the observers. "If we were at the galactic center, the effect would be obvious," Guthrie says. Moreover, he adds, newer, still unpublished data seem to confirm the effect.

Guthrie and Napier's analysis has impressed some observers with its thoroughness. But most astrophysicists want to see light-tight evidence before they start to rewrite physics. "They are examining an interesting question, and they have done a very good job," says P. James E. Peebles, an astrophysicist at Princeton University. But, he cautions, "they have not produced manifest evidence" for redshift quantization.

Peebles suggests that some biases in the data might persist despite Guthrie and Napier's best efforts. If some of the galaxies in their sample were unrecognized members of large clusters, for example, spurious periodicities might arise. "It is interesting but only mildly interesting—I might think about dragooning a graduate student into looking at it," Peebles concludes.

Edwin E. Salpeter, an astrophysicist at Cornell University, says he and others saw a surfeit of redshift differences of around 70 km/sec between members of twin galaxies a few years ago—apparent support for Tifft's original idea. But Salpeter became convinced the effect was a result of the way candidate pairs of galaxies were selected. With improved selection procedures, he says, typical redshift differences between members of a pair became smaller, as he had expected.

In addition, Salpeter and some other astrophysicists are troubled by Tifft's tendency to keep modifying the quantized redshift hypothesis. He notes that Tifft introduced fractions of 72 km/sec when new data seemed to demand it. "You can't change the law of quantization when the data change," he asserts.

Tifft has recently extended the redshift quantization hypothesis even more radically to include the notion that redshifts change over time. Moreover, he holds, they change in a way that depends on where they fall in relation to the periodicity. Redshifts might be oscillating, Tifft suggests. Although hampered by a scarcity of data, Tifft has elaborated a complex theory based on a 2.6657-km/sec redshift quantization, with different multiples dominating for different types of galaxies.

Nevertheless, Tifft and Cocke have their supporters. Among them is Peter A. Sturrock, director of the Center for Space Science and Astrophysics at Stanford University, who says they should be given a fair hearing. "It isn't something astronomers should brush off," he insists. Sturrock thinks the question of redshift quantization could be settled definitively in a collaborative project. But he says the participants would have to specify in advance what kind of observations would disprove the theory.

Tifft has no fears: he is eagerly awaiting accurate data from the 300-foot radio telescope now under construction at the National Radio Astronomy Observatory in Green Bank, W.Va. The results will, he maintains, either confirm or disprove redshift quantization for once and for all. —*Tim Beardsley*

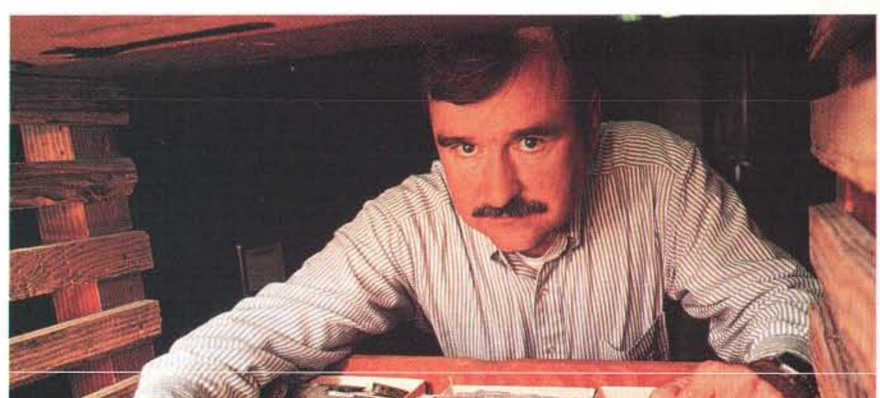
Cambrian Jolt

How often did mass extinctions reshape evolution?

Several times in the history of life, something has given the global ecosystem such a jolt that a large fraction of all species became extinct within a few million years—the blink of an eye in the vastness of geologic time. Paleontologists generally recognize five events on that scale in the past 500 million years. The most recent—and famous—spelled the end of the dinosaurs, about 65 million years ago.

In mid-1992 Philip W. Signor, a researcher at the University of California at Davis, made paleontologists sit up and take notice when he announced the discovery of a sixth mass extinction, one he asserted was more extensive than any of the others. Signor's event occurred early in the history of life, more than 500 million years ago, in the early Cambrian period. Multicellular animals were then only just starting to swim in the seas. A mass extinction at that time might, paradoxically, help to explain the so-called Cambrian explosion—the relatively rapid emergence of many different animals during the early Cambrian.

Other mass extinctions are known to have spurred the appearance of new classes of organisms by clearing the ecological stage. Mammals, for example, proliferated after the demise of the dinosaurs. If Signor is right, catastrophe might have been part of life's story from the start. "It argues for the importance



STEPHANIE RAUSSER

PHILIP W. SIGNOR of the University of California at Davis believes he has found fossil evidence of the earliest, most extensive mass extinction.

of contingency," Signor says. "Things broke the right way for mollusks, but it might have been otherwise. Just as if Joshua Chamberlain hadn't been at Gettysburg and hadn't told his men to fix their bayonets, that battle could have gone differently. It wasn't the inherent superiority of the North."

The previously unrecognized mass extinction occurred, Signor believes, about 10 million years after the start of the Cambrian period, 510 to 560 million years ago, causing about 80 percent of all living genera to become extinct over a period of five million years. The Cambrian casualty list includes echinoderms (the ancestors of today's sea stars and starfish), brachiopods (often called lampshells) and true mollusks.

In contrast, 60 percent of all genera went extinct within five million years during the mass extinction at the end of the Permian period 300 million years later. That extinction has until now been considered the biggest, so on a raw percentage basis Signor's event would be the biggest as well as the oldest. Signor bases his conclusion on a data base of 850 Cambrian genera that he assembled from other investigators' reports.

Although other workers had previously noticed some disruption during the Cambrian, Signor suggests they did not notice the scale of the event because much of the recent research on early Cambrian animals had not yet been catalogued. Françoise Debrenne of the Institute of Paleontology in Paris, for example, convinced her colleagues years ago that many of the primitive reef-forming sponges called archaeocyathans disappeared then. And John J. Sepkoski, Jr., professor of paleontology at the University of Chicago and an authority on mass extinctions, has also written about large-scale extinction in the early Cambrian. But he acknowledges that Signor has assembled more data and has improved the accuracy of the timing.

Yet Signor's colleagues tend to be skeptical. "Signor's data are difficult to interpret because the early Cambrian was a time of very rapid turnover," Sepkoski says, and multicellular animals "were just getting established." Against that turbulent backdrop, when roughly half of all genera routinely disappeared every five million years, the extra surge of extinctions that Signor has documented corresponds to less than a doubling of the rate. In contrast, the event that finished off the dinosaurs—which many researchers think was triggered by a meteorite or massive volcanism—represented a 10-fold increase in extinction rates, Sepkoski points out.

Other scholars are not likely to accept Signor's ideas until he produces a complete report of his findings. So far he has described his conclusion only in a one-page summary. He also admits he has no idea what caused the event—the geologic record offers no evidence of a meteoritic impact or an eruption, for example. But the sea level apparently fell during the early Cambrian, perhaps as a result of plate tectonic activity and a shift in ocean currents. Even a small temperature change can make marine organisms less fertile, precipitating their extinction, notes Michael E. Taylor of the U.S. Geological Survey in Denver.

Allison R. Palmer, head of the Institute for Cambrian Studies in Boulder, Colo., and an expert on trilobites, says he is not convinced that Signor's early Cambrian event is anything more than an artifact. Palmer has shown there were several continental-scale extinction events later in the Cambrian—although nothing on the global scale Signor is proposing. Palmer questions the accuracy of Signor's calendar, observing that trilobites provide no evidence of extraordinary disruption at Signor's crucial time. Simon Conway Morris of the University of Cambridge, who studies other Cambrian fossils, is also cautious about accepting Signor's claim. Understanding of the taxonomy of early Cambrian creatures "is in pretty poor shape," he points out. Imperfect taxonomy can bias conclusions that are based on counting groups of animals.

"It's not credible to me that there's a tremendous bias from the taxonomy," Signor fires back. And even if his timing of the event is not precise, he argues, "that shouldn't obscure the fact there was a major reduction in diversity" of animals early in the Cambrian. Signor, though admitting to being a "little ashamed" that he has not gotten around to writing up his big claim in a formal paper, has given paleontologists an idea that will probably be discussed for some years to come. —*Tim Beardsley*

Cellular Response

Are antibodies the most effective defense against AIDS?

Over a decade into the AIDS epidemic, more than a dozen candidate vaccines that might curtail the spread of the disease are in preliminary testing. Whether they will be effective at protecting against infection in humans remains to be seen. Nearly all are aimed at provoking an antibody response to the human immunodeficiency virus (HIV), the causative agent of AIDS. New experiments suggest that some people, sometimes, may be able to resist HIV infection. But their ability to fend off the AIDS virus may depend on an entirely separate arm of the immune system, known as cellular immunity.

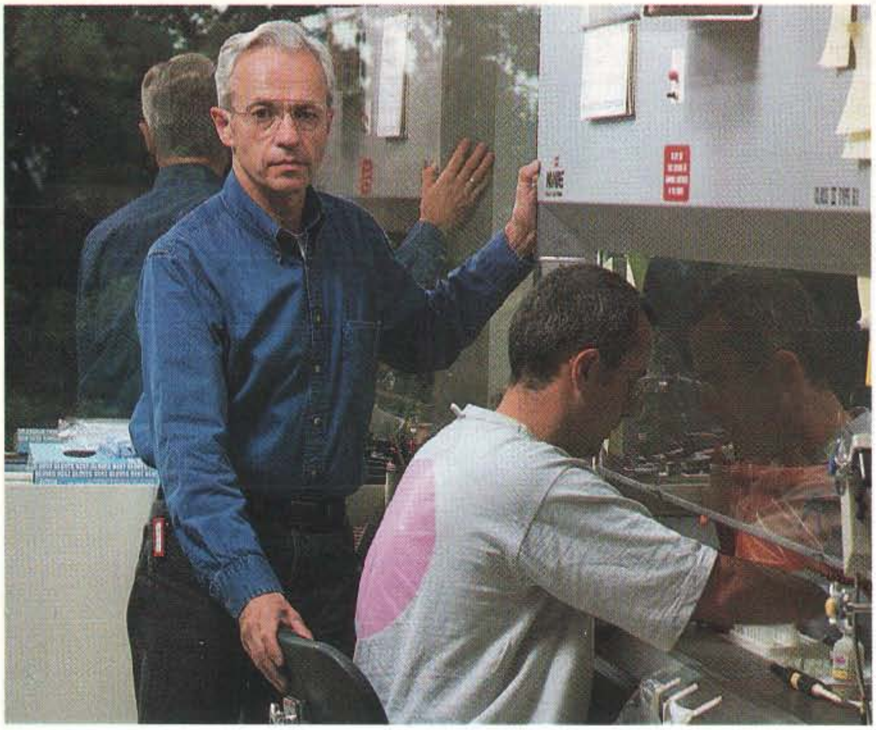
Key observations have been made by Gene M. Shearer and Mario Clerici of the National Cancer Institute. In an article published in the *Journal of Infectious Diseases,* they have shown that in some people *T* lymphocytes have responded to components of HIV even though these individuals have no anti-HIV antibodies and are apparently uninfected. Most of those individuals have had potential exposures to HIV. Shearer and Clerici believe that in such cases the *T* cells have successfully fought off the infection.

Indeed, the antibody-producing cells in HIV-positive patients who have a well-established infection seem to be overstimulated. But the cells that are responsible for cellular immunity are impaired. Shearer and Clerici interpret that and other observations as suggesting that only if cell-mediated immunity fails will the immune system start to produce antibodies—which are ultimately unsuccessful at containing the virus.

If Shearer and Clerici are right in believing that there is a successful natural cellular immune response to HIV, candidate vaccines could be sought that stimulate the cellular part of the immune system. Whether or not a vaccine stimulates the production of antibodies might be less important. The researchers even venture the idea that stimulating antibodies might be harmful, because it could impair the cellular part of immunity. They base their opinion on studies indicating that the two arms of the immune system might actually inhibit each other.

Although Shearer and Clerici's work has attracted notice, researchers are not yet ready to forget about antibodies. "It seems to me very plausible that there are people who have been exposed to the virus but in whom infection was not established," says Daniel F. Hoth, director of the AIDS program at the National Institute of Allergy and Infectious Diseases (NIAID). "I accept that Shearer and Clerici can detect the immunologic memory of exposure. But it doesn't tell you if the cellular response is functional. It may be just a partial immune response to an abortive infection."

Most of the AIDS vaccines now being



JOHN MCGRAIL

IMMUNE SYSTEM might be able to defeat the AIDS virus, say Gene M. Shearer and Mario Clerici of the National Cancer Institute.

tested for safety in healthy volunteers are based on pure proteins, which stimulate the production of antibodies. Some animal studies have found that a strong vaccine-induced antibody response can indeed protect chimpanzees against HIV. But it is still unclear whether an adequate level of antibodies can be maintained or whether the effect will work in humans.

Margaret I. Johnston, associate director of the NIAID program, points out that even in vaccines that confer effective protection against other diseases, the central components responsible for immunity are rarely known. Historically, she says, successful vaccines have usually used either live attenuated viruses or mixtures of components, not just pure proteins. "My bias is that an effective vaccine will need to produce a cellular response as well as antibodies," Johnston says.

Some vaccine developers apparently agree: several new candidate products are designed to stimulate cellular immunity as well. Some, such as those produced by Pasteur-Merieux in Lyons, France, and by Wyeth-Ayerst International in Philadelphia, utilize viruses modified to include components of HIV. These products should induce cellular responses as well as antibodies, Johnston notes. (Bristol-Myers Squibb was conducting clinical trials of a HIV vaccine based on an engineered vaccinia virus—the same one used in smallpox vaccine—but discontinued them a year ago because of safety fears.)

Jonas Salk, who developed a successful polio vaccine in the 1950s and is investigating a vaccine made from killed HIV, has embraced Shearer and Clerici's ideas about cellular immunity. Many researchers have concerns about the safety of a vaccine made from the whole virus, however. Others are following a more cautious approach they hope will nevertheless produce cellular responses. Therion Biologics in Cambridge, Mass., for example, is investigating the potential of "pseudovirions"—particles formed from HIV proteins that may be able to induce a better immune response than isolated proteins.

Cells in the mucosal surfaces may be among the first to become infected in many instances of HIV transmission. If it becomes clear that immunity to HIV can be induced, vaccine developers will have to ensure that their products are protective against sexual exposure as well as intravenous injection, Johnston points out. Some experiments indicate that research primates can be immune to the simian HIV-like virus, known as SIV, when challenged by injection but not when challenged vaginally. Yet sexual immunity is possible, at least for SIV in macaques. Preston Marx of the New Mexico Regional Primate Research Laboratory has shown that macaques can be made immune to vaginal challenge using a vaccine made from whole killed virus particles.

Johnston has no doubt that vaccine development has to forge ahead even though important questions about the immune response to HIV remain, for now, unanswered. "Our strategy is to follow a very directed critical path," she says. Phase 2 trials of vaccines in the U.S.—to evaluate safety in individuals at risk—are planned to start within the next four months. "We have to be reasonably cautious and move ahead as fast as we can," Johnston says. "None of us feel we can sit around and wait for all the answers." *—Tim Beardsley*

Thinking Green

Can environmentalism be a strategic advantage?

Robert Ferrone doesn't have to think hard to remember a time when fellow engineers eyed him with skepticism. As leader of environmental efforts at Digital Equipment Corporation (DEC), Ferrone's presentations on "environmentalism as a strategic technology" left most of his colleagues cold. "They'd come up to me afterward and say, 'I don't see anything strategic about the environment,'" he recalls.

That was last year. This year public concern has loaded Ferrone's message with urgency. Now, he reports, "people keep coming up to me saying, 'The environment is important. I haven't internalized it all yet, but I understand the nature of what you're saying.'"

Minding the environment is a complicated business. Recycling discards and reducing the waste created during manufacturing are becoming corporate doctrine. More radical—but far more promising, advocates argue—is "green design" or "design for the environment" (DFE). "We've been dealing with the symptoms of environmental problems," says Braden R. Allenby, a visiting fellow at the National Academy of Engineering and a champion of DFE. In contrast, he notes, DFE calls for "fundamental shifts in the way we manage technology."

Like long-standing industrial programs that support efforts to "design for quality" or "design for manufacturability," DFE seeks to stir engineers to think about the environmental implications of a product and of its manufacture during the earliest phases of design. These considerations may em-

brace a sweeping collection of issues: the environmental distress caused by obtaining the raw materials, the toxicity of using and discarding chemicals during production, the likelihood that the product itself can be refurbished, reused or at least recycled once the consumer has decided to abandon it. "I call it cradle-to-cradle design," Ferrone says.

The concept seems to be gaining currency in Washington. A recent report from the Office of Technology Assessment suggests that government consider using legislation and other means to encourage green-design efforts. Agencies, including the military, are beginning to include issues of pollution and waste in their procurement decisions.

Particularly in the electronics industry, which is a relative neophyte to such concerns, U.S. companies are still a far paler shade of green than their overseas competitors. "Environmental technology will be a rite of entry internationally in the next decade," declares Brian G. Kushner, a vice president at Microelectronics and Computer Technology Corporation (MCC).

Many European firms are preparing for the "take back" laws that countries, starting with Germany, are putting in place. Japanese manufacturers are seeking larger market share by virtue of their "greenhood." They also are selling cleanup and control technologies. Although attitudes are far from unanimous, "we're seeing environmental issues shift from being a cause celebré of environmentalists to being a business issue," Kushner notes.

A handful of U.S. corporations are winning green points. DEC and Xerox, for instance, accept old equipment from customers; in some cases, Xerox also refurbishes and resells the machinery. By the end of 1992, IBM was slated to announce customers could return any of its products. These and other companies are also beginning to label components by type of material to facilitate recycling and to investigate techniques for disassembling and reusing old products. A dozen computer makers and chip producer Intel have said they intend to earn a "green star" from the Environmental Protection Agency by tuning their products so that they consume significantly less energy when idle.

Nevertheless, rhetoric and good intentions still outweigh substantive changes. Kushner and other green-design advocates argue that treating the pollutants at the end of a factory exhaust pipe is not enough. "Everyone positions this as a manufacturing issue," Ferrone says. "It's not. It's really a design problem."

But few electronics makers are shouldering the arduous task of redesign. At DEC, engineers are talking about replacing some of the nuts and bolts that hold together computer casings with plastic tabs. Even this kind of change will not happen soon, Ferrone concedes. "No one should be misled into thinking this is a smooth transition," Allenby warns. "No one has yet figured out how to translate environmental data into the systems analysis information that design teams can use."

Electronics firms are still in the early days of analyzing which materials may be particularly harmful to the environment—or to workers. And even when likely risks are spotted, manufacturers are slow to change their processes. For instance, although a recent IBM study seems to support the implications of a two-year-old report funded by DEC that some standard chemicals used in semiconductor manufacturing may provoke miscarriages in workers, chip makers are not yet abandoning the solvents. Another study on the toxicity of these chemicals, undertaken by researchers at the University of California at Davis, was slated to be released in December.

Meanwhile the government is stepping up the pressure. The Resource Conservation and Recovery Act of 1976, which regulates waste disposal, is up for reauthorization. Congress continues to show an appetite for fresh environmental legislation. In early October a committee recommended that Sematech, the chip research consortium, devote 10 percent—or $10 million—of its 1993 budget to environmental issues. (Sematech representatives maintain they are already exceeding that mark.)

To try to get a jump on future regulations, MCC has been leading an effort to outline the environmental and pollution costs of computers throughout their life cycle. Ferrone hopes the report will establish an environmental agenda for the industry—complete with quantitative milestones for measuring progress. "Otherwise, it's just rhetoric," he says.

Electronics manufacturers are also looking to the government for aid. The national laboratories could be charged with exploring environmental manufacturing issues, some suggest. Others point with envy to funding and tax breaks that European and Japanese companies receive for green investments.

Undertaking environmental design and new processing techniques will be costly, Ferrone admits. But disposing of equipment and chemicals may be more so. "The priority is profitability," he states. Now electronics manufacturers must begin to count their costs—and benefits. *—Elizabeth Corcoran*

Running Hot

Will jet engines create a market for advanced materials?

Turbine blades in jet engines are exposed to temperatures of more than 2,500 degrees Fahrenheit—a hundred or so degrees from turning them into the metallic equivalent of peanut butter. When U.S. Air Force materials and engine specialists suggested 10 years ago that a new set of materials was needed that would be capable of withstanding temperatures as high as 4,000 °F, engine makers scoffed. "I can remember thinking, 'These guys are crazy,'" says Lyman A. Johnson, a materials manager with GE Aircraft Engines.

The program's larger goal of doubling performance for engines in jet fighters, helicopters, transport aircraft, missiles and perhaps ultimately civilian aircraft has yet to be achieved. But the joint industry-government program that has emerged has become a testbed for the most advanced "designer" materials based on metals, polymers and ceramics.

The Integrated High-Performance Turbine Engine Technology (IHPTET) initiative is a counterpoint to the much repeated claim that U.S. industry cannot look beyond the next 90 days. Begun five years ago, the program extends through 2003; total spending by that time could reach $3.5 billion. IHPTET combines the efforts of the U.S. Army, Navy and Air Force and the seven domestic gas-turbine engine manufacturers, as well as the National Aeronautics and Space Administration and the Defense Advanced Research Projects Agency.

IHPTET, which also focuses on developing new aerodynamics and engine designs, will provide an extended test of whether military aerospace technology can help commercial manufacturers. "You can take the core of a military engine, put a new fan and turbine on it and make it into a commercial engine," says Ralph E. Anderson, group leader of light metals and processing for Pratt & Whitney.

The jet engine is already a high-performance machine. The thermal efficiency of the gas turbine—its ability to convert the energy in the fuel into useful work—currently peaks at about 33 percent, approximately eight percentage points higher than the internal-combustion engine in an automobile, according

to Eugene E. Covert, a professor of aeronautics at the Massachusetts Institute of Technology. Advanced metals and ceramics, which allow an engine to run at higher temperatures and thus produce better fuel combustion, will raise the efficiency rating to about 40 percent.

The turn-of-the-century jet engine, though smaller, will look similar to the machine in common use for the past 50 years, says James S. Petty, the air force's IHPTET program manager. But inspection of the post-2000 engine with an electron microscope would reveal big differences. Critical parts of the engine now made of nickel or titanium superalloys will be replaced by composites in which particles or fibers are infused into a matrix of polymer, metal or ceramic. The size and placement of the fibers will allow a materials engineer to adjust tensile strength and other parameters precisely—the reason for the designation "designer" materials.

But perhaps IHPTET's most important success in materials stems from advances in metal composites and intermetallic compounds. The Allison Gas Turbine division of General Motors achieved a milestone a year ago by making the first rotating part, a compressor ring, from a titanium alloy embedded with silicon carbide fibers. The thin ring, which holds compressor blades on the engine shaft, weighs 10 pounds and replaces a nearly 60-pound solid disk. Because metal-matrix composites are stronger than the alloy parts, fewer rotors are needed. "You can get a rotor that runs faster and one that is able to achieve higher pressures," says Ronald E. York, Allison's director for advanced engine projects.

Allison made the ring by interleaving helical coils of titanium with others that contain the woven fibers. The coils are then squeezed together at high temperatures and pressures, a process York compares with fusing Slinkies that are sandwiched together. Designing the coils severely challenged Allison materials and structural engineers: each fiber has to be carefully oriented within the metal matrix to provide the maximum resistance to centrifugal forces generated as the rotor turns. Allison intends to adapt this technology to a planned jet-fighter engine, although GM is now trying to sell this division. Other engine manufacturers, including GE and Pratt & Whitney, plan to test their own versions of composite rings within a year or so.

If the rotor is to obtain the maximum speeds and withstand temperatures of as much as 1,400 °F, the titanium alloy must be replaced by an intermetallic alloy. In intermetallics, different atoms—titanium and aluminum, for example—become interspersed with one another within the crystal lattice. In doing so, they form an ordered arrangement that, in some cases, alternates a layer of titanium with one of aluminum. Some of the titanium aluminides under development will be able to survive at temperatures 300 to 400 degrees hotter than those withstood by existing titanium alloy rotors.

In the ultimate rotor the silicon-carbide reinforcing fibers used to fabricate current rotor prototypes may be replaced by those made of alumina. Silicon-carbide fibers expand at a rate different from that of the titanium-aluminide matrix, so they can cause cracking of the composite material.

As IHPTET progresses, engine manufacturers will confront the difficult question of whether these materials can be manufactured at a cost that warrants putting them in production engines—an issue made more acute by the lagging market for military and commercial aircraft. According to York, a rule of thumb states that no more than $5,000 should be spent for each pound of weight saved—a figure that Allison is not close to meeting for its metal composites.

Today assembly lines for metal-matrix composites do not exist. "You tend to have Ph.D.'s standing around making these things up by hand," says Walter H. Reimann, chief of the Materials Development Branch at Wright-Patterson Air Force Base. Help may be on the way. With funding from a separate air force program, Textron, a conglomerate with substantial aerospace interests, is setting up a pilot facility to manufacture metal-matrix composites.

Over time, commercial demand could bring the costs down for the aerospace industry. GM, Ford Motor and a number of Japanese automobile manufacturers are testing one titanium aluminide for lightweight, high-temperature engine valves. But don't wait for golf clubs or bicycles made from a metal-matrix composite. "We don't have a consumer item that could be made in large quantities as a toy for wealthy people," says GE's Johnson. *—Gary Stix*

Finding Fault

Can seismologists predict earthquakes?

In mid-October a moderate earthquake, measuring 4.7 on the Richter scale, rumbled into the tiny village of Parkfield in central California. The town was outfitted with creep meters, laser triangulation devices, and millions of dollars' worth of other seismological equipment that had been patiently waiting in ambush for such an event to occur for the past seven years.

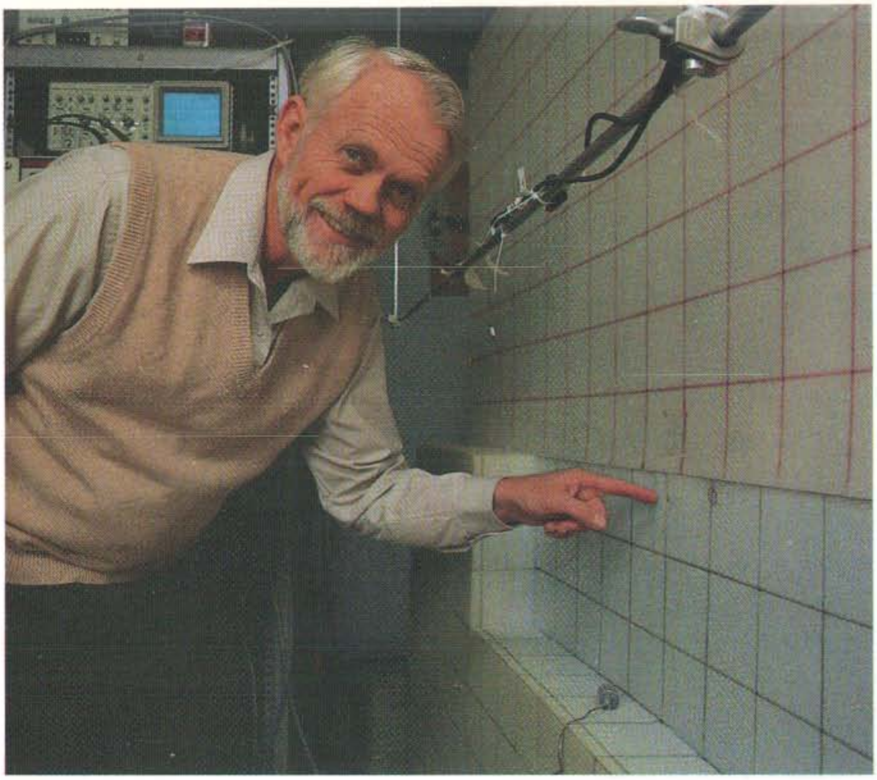
Parkfield is the staging ground for the firmest forecast ever made by seismologists from academe and the U.S. Geological Survey: the prediction that a magnitude 6 earthquake would occur by the end of 1992. This assertion was based on the relatively regular recurrence there of large earthquakes at 22-year intervals dating back to the 19th century.

With barely a few weeks left, the San Andreas fault granted the seismologists part of their wish. And USGS scientists initiated a vigil—a "level A" warning—with hopes that more was to come. "We're all fervently praying that the damn thing gets its act together," says Allan G. Lindh of the USGS's Menlo Park office.

This half-fulfilled wish of the scientists was a mocking taunt from the famed and feared San Andreas fault. It also concisely summarizes the turmoil that exists about the regularity of seismic activity—and ultimately the ability to predict earthquakes. The Parkfield experiment is based on a model of how earthquakes occur that scientists have relied on in one form or another for about 100 years. The seismic-gap theory, as it is called, proposes that stress builds up on each section of a fault over a period of time. A quiescent segment of the fault, the location of a "gap" in seismic activity, then suddenly gives way in an earthquake, releasing the strain energy in the rocks. The process then repeats itself.

But that model was called into question in a controversial paper published in the *Journal of Geophysical Research* in December 1991. In the article Yan Y. Kagan and David D. Jackson of the University of California at Los Angeles reported that a well-known attempt to apply the theory to fault zones in the Pacific rim had failed to forecast the location of 40 large earthquakes between 1979 and 1989. Other seismologists have begun to question whether apparently repeated events, like those in Parkfield, actually follow an identical pattern, since they may have occurred on different fault zones.

Two authors of one of the original seismic-gap forecast studies, Stuart P. Nishenko of the USGS and Lynn R.


VAN YATES *Black Star*

***JAMES N. BRUNE** of the University of Nevada at Reno believes foam blocks may tell more about the dynamics of earthquakes than do rocks from an actual fault zone.*

Sykes of Columbia University, have submitted a yet to be published rebuttal to the *Journal of Geophysical Research.* They assert that half of the earthquakes Kagan and Jackson examined did not meet the criteria (for size, type of plate boundary and so on) they had set out for their original predictions.

The debate does not end there. Seismologists have been wrestling with the seeming anomaly that the San Andreas fault is experiencing high levels of seismic activity but that the levels of stress along the fault are rather weak—and thus heat from sliding of the rocks under high frictional stress is absent.

A number of other ideas have suggested that the way faults rupture may be subject to chaotic phenomena that complicate earthquake prediction based on models of regularly repeating build-up and release of strain in the rock. Theories proposed separately by seismologists James N. Brune of the University of Nevada at Reno and Thomas H. Heaton of the USGS in Pasadena suggest that, once initiated, an earthquake may ripple in a wavelike motion down the length of the fault, decreasing the pressure with which the two opposing sides of the fault are held together.

This effect can cause a large motion of the two sides of the fault, even in the absence of large amounts of stress. Brune, who found confirmation for this rippling effect while working with foam blocks, compares it to the carpet installer's trick of moving a wrinkle to adjust a rug's position, an alternative to the impossible task of sliding the entire surface of the carpet.

In Brune's hypothesis, the magnitude of any earthquake is proportional to the uncertain length that this rippling motion propagates down the fault. The progression of this movement along the fault may continue until it is stopped by rocks rigidly held together or by a bend in the fault. But where and how the movement stops may be impossible to know, Heaton claims. He believes the difference between a small earthquake (there are 40 every day in southern California) and the much feared "Big One" depends on what he calls "rupture dynamics." Heaton suggests that a "chaotic" feedback interaction can amplify slippage of the fault. The amount of friction and stress controls the amount of movement of the fault, while, conversely, the motion of the rupture continuously changes the amount of friction on the fault.

Some of these new ideas make life difficult for earthquake forecasters. An unstable, chaotic system means that large stresses may sometimes fail to trigger a big earthquake, whereas at other times small events may set off a mighty shaking. The interval between earthquakes may be either short or long.

Others, such as Lindh, are not ready to part with the old model, pointing to apparent successful predictions, such as the Loma Prieta earthquake, which caused extensive damage in the San Francisco Bay Area. Lindh believes some of the newer theories, though untested, may complement the standard view, explaining the minority of earthquakes that do not fit the conventional mold. Lindh characterizes Brune's and Heaton's ideas as "interesting but highly speculative."

Lindh contends that over the next few decades enough pieces of the seismic puzzle will fall in place to form a clearer picture about when and where the next Big One is likely to occur. For the moment, though, even he is cautious. He invokes the words of Charles Richter, the California Institute of Technology seismologist whose name is on the magnitude scale: "Only fools, liars and charlatans predict earthquakes." *—Gary Stix*

Unlikely Messengers

How do nerve cells communicate?

Two compounds—one playing the role of a king, the other, an anarchist—have offered dramatically different insights into the governance of neurotransmission. The first, a form of cellular energy called adenosine 5′-triphosphate (ATP), expands the classical understanding of the substances that convey information between nerve cells and may offer a glimpse into the pedigree of the nervous system. The second, nitric oxide, commonly known as a toxic air pollutant, challenges the status quo and suggests a novel means of neuronal communication.

Despite their differences, both compounds seemed to many scientists unlikely messengers. For that reason, the identification of ATP as a significant neurotransmitter and the investigation of the neuronal action of nitric oxide have opened new avenues for research and treatment. "Both reshape our thinking in slightly different ways," says Floyd E. Bloom of the Scripps Research Institute in La Jolla, Calif.

The discovery that ATP is a neurotransmitter is a vindication for Geoffrey Burnstock of University College, London. His theory of ATP-based neurotransmis-

sion, proposed in the 1970s, was not quietly ignored; it was ostracized and ridiculed. "At one stage I was very hurt," Burnstock recalls. "One researcher said to me, 'I am going to devote my life to destroying your hypothesis.'"

Although it has been known for several decades that ATP is stored in synaptic vesicles—neuronal containers for neurotransmitters—in the central nervous system, it was thought to be a mere flunky, an aid, to such messengers as noradrenaline and acetylcholine. "ATP is like fresh air—it is so ubiquitous," observes Eric R. Kandel of Columbia University. "It just didn't seem to have the characteristics of a neurotransmitter. We were all blind."

But two recent studies have finally provided the requisite evidence that ATP is indeed a major neurotransmitter. Richard J. Evans, Victor Derkach and Annmarie Surprenant of Oregon Health Sciences University reported in *Nature* that ATP works as a neurotransmitter in the peripheral nervous system. Using a drug called suramin, which blocks receptors to ATP, they were able to thwart neurotransmission in cell cultures. The researchers concluded that ATP was responsible for the undelivered message.

Several months later ATP was found to reign in an area of the central nervous system as well. Frances A. Edwards, Alasdair J. Gibb and David Colquhoun of University College, London, also reported in *Nature* that preventing ATP from binding with its receptors blocked neurotransmission in a region of the rat brain called the habenula.

The studies suggest that ATP is not merely a neurotransmitter but that it is one of the elite. Until it was recognized, only two substances were known to produce a fast—that is, within milliseconds instead of within several seconds—as well as excitatory response: glutamate and, at certain receptors, acetylcholine. "To find a new one is a kind of excitement just in itself," Surprenant says. "It has been years and years since a new fast guy has come along."

Once the five proposed ATP receptors have been cloned and drugs that bind more specifically with them are developed, researchers expect to map out the pathways in which ATP is active in the brain—and possibly design drugs to treat disorders of the nervous system. In addition, researchers are looking to ATP for clues about the evolution of the brain. "It is probably the most primitive extracellular messenger of them all," Burnstock explains, adding that there are ATP receptors in amoebas, insects and fish.

Researchers are also intrigued by the unusual action of nitric oxide. Unlike other neurotransmitters, it is toxic, and it is not stored in vesicles but rather is produced on demand. And most unbefitting a neurotransmitter, but appropriate for an anarchist, it diffuses directly through cell membranes to bind with a peculiar receptor: iron. "Nitric oxide does not change the way we think about classical neurotransmission, but it expands the way we think nerve cells signal each other," comments Zach W. Hall of the University of California at San Francisco. "The signaling molecule is always thought to act on the outside of the cell, which then tickles the insides. But here is a case where the messenger zooms across the membrane."

Interest in nitric oxide also stems from a search for the chemical basis of memory, a phenomenon called long-term potentiation (LTP). In this model of memory, communication between certain nerve cells is strengthened, creating a record of neuronal communication that may serve as the foundation for long-term memory.

Researchers have postulated a yet to be found renegade compound—one that travels backward from the postsynaptic neuron (the receiver) to the presynaptic neuron (the sender), enhancing the relation between them. There are several candidates for such a retrograde messenger, including arachidonic acid and nitric oxide. "The modern history of neurobiology is based on the idea that information flows in one direction," Kandel says. "The interest in nitric oxide lies not just in the initial discovery but in the fact that it can flow in a backward direction."

Four teams of researchers have described how they prevented LTP by blocking the production of nitric oxide. But for now, this possible facet of nitric oxide's role in the brain remains unclear. "The case for nitric oxide is unproved at best," remarks Timothy V. P. Bliss of the National Institute for Medical Research in London, who first described LTP some 20 years ago. For example, Bliss finds that interfering with nitric oxide production does not prevent LTP from taking place at body temperature, only at room temperature.

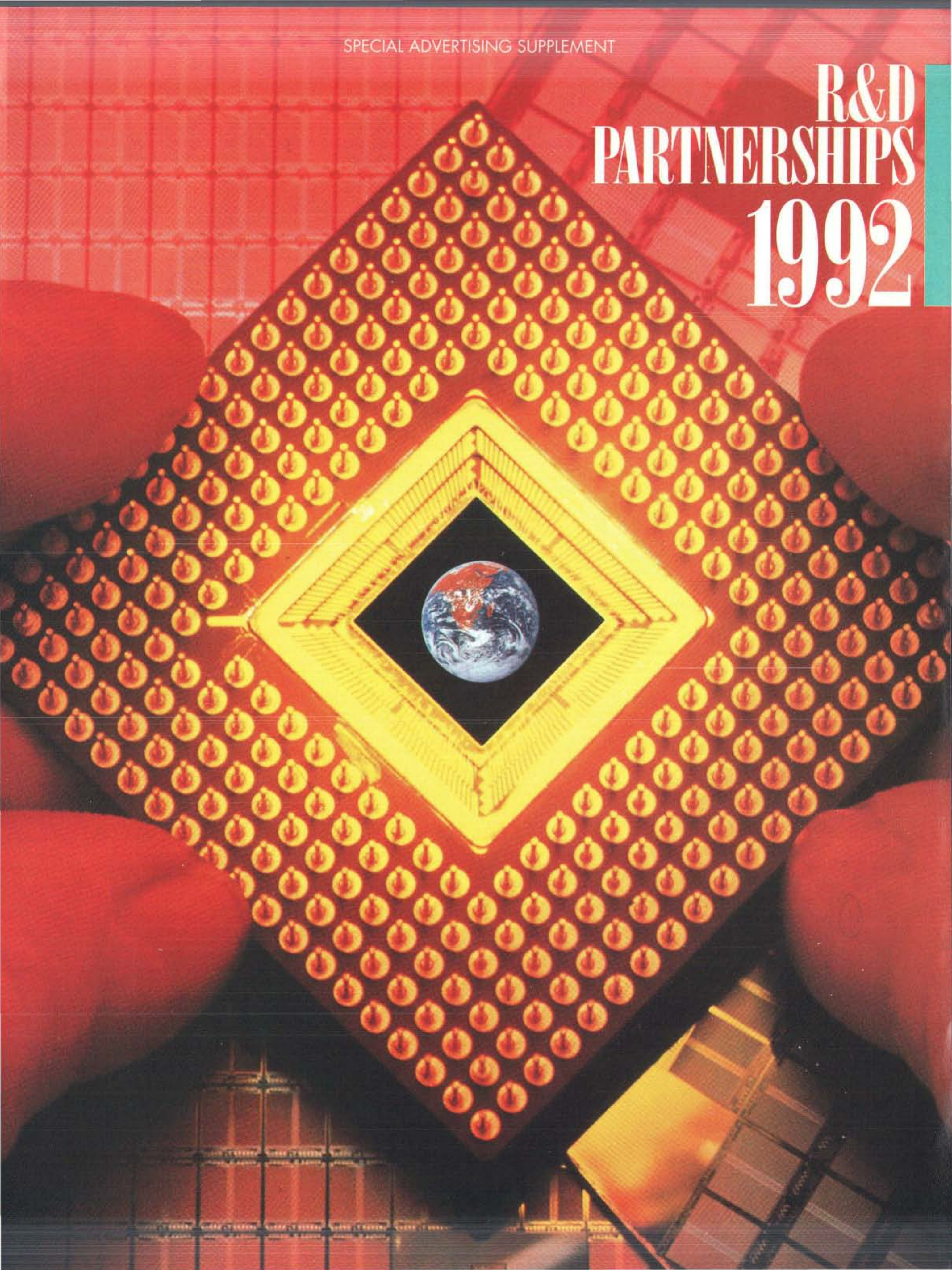
The search for neurotransmitters is far from over. One candidate is carbon monoxide. Solomon H. Snyder of Johns Hopkins University and his colleagues have data indicating that carbon monoxide affects the same second messengers—intracellular relay compounds that pick up the baton once a neurotransmitter binds with a receptor—that nitric oxide does. And "maybe there are other bizarre, unstable, weird chemicals that might be analogous," Snyder says. —*Marguerite Holloway*

KRISTIN FINNEGAN

NEW FAST GUY of neurotransmission is ATP, explains Annmarie Surprenant of Oregon Health Sciences University, whose team studies the chemical's novel role.

SPECIAL ADVERTISING SUPPLEMENT

R&D PARTNERSHIPS 1992



C. William Gear, President, NEC Research Institute

THE PURSUIT OF PURE RESEARCH

Look into the eye of a fly and you might find the future of computers and communications. Sit down and play chess with a computer and ponder how to make machines that can think intuitively. Study Noam Chomsky to uncover the innate principles that may enable a computer to understand any human language.

Necessity is not necessarily the mother of invention at the NEC Research Institute, Inc., in Princeton, N.J. A risk-taking, unorthodox approach to fundamental research is not only accepted but encouraged at the two-and-a-half-year-old, $22 million research facility that operates as a wholly owned subsidiary of NEC Corp., the $28 billion global electronics company headquartered in Tokyo. NECI, as it is known, is the only pure unrestricted, undirected research institute sponsored by industry today. Structured in the tradition of AT&T Bell Labs and Watson Research of yesteryear, NECI is dedicated to free-form research in computer and physical sciences as it relates to the processing and interpretation of information.

"We don't have projects," explains C. William Gear, the newly named president of NECI. "We hire outstanding people who are doing interesting research that will have consequence in the future." To date, there are 18 Ph.D.s in the institute's physical sciences and computer sciences divisions. NECI's ultimate goal is to have up to 27 scientists in each division. Some of the outstanding researchers include Robert Tarjan, winner of the Turing Award in Computer Sciences and a leading expert in algorithms; Eric Baum, the physicist turned computer scientist from Princeton, whose expertise lies in computational learning theory and game research; William Bialek, a biophysicist from Berkeley, who is analyzing how a fly's eye works, and Stuart Solin, an expert in materials science and a former director of the Michigan State University Center for Fundamental Materials Research.

For scientists, NECI has strong appeal. "Stable funding in long-term, fundamental research is hard to find," Gear says, noting that today, industry is forced to quickly commercialize its research findings and academia can no longer attract sufficient long-term funding. NEC Corp. requires only that research pertains to its twin interests of computers and communications. A board of the most senior scientists oversees scientific direction.

A RISK-TAKING UNORTHODOX APPROACH TO FUNDAMENTAL RESEARCH IS ACCEPTED— AND ENCOURAGED— AT NECI.

In the physical sciences division, scientists pursue how humans and other biological systems handle information. The eye of a fly, for instance, is capable of processing information at a precision that approaches the limits imposed by physical laws. By discovering how biological systems work at this level, scientists hope to find applications in computation. Research also involves the examination of alternative materials and processes to overcome current physical limitations in system components.

The computer sciences division is organized around four key areas most likely to influence how computers work: architecture, interface, intelligence, and theory. The work of each subdivision enhances the value of the others. In architecture, for instance, a goal is to overcome the obstacles faced in parallel processing so that unlimited amounts of parallelism can be harnessed at one time. Increased parallelism must be accompanied by improved user interface; therefore, interface research becomes critical as well.

NEC Corp. provides NECI's current annual budget of $20 million and aggressively pursues a free interchange of information with the global scientific community. NEC and NECI co-sponsor an annual symposium on computer and communications research, which is held alternately in Princeton and Japan. The symposium this year focused on computational learning theory, machine learning and cognition, and drew an audience of almost 200 to Princeton. Every summer, the institute sponsors a week-long series of lectures in biophysics which is open to all U.S. university faculty and graduate students. And in the interest of fostering community relations and encouraging the choice of science as a career, NECI hosts a two-week summer course for inner-city children from nearby Trenton.

The fruit of NECI's efforts may not be evident for years. The very nature of long-term research requires great patience, an attribute that both NECI and NEC Corp. share. Dr. Gear notes that the NEC Corp.'s motto is "computers and communications for human potential." NECI's unorthodox approach to that potential may produce breakthroughs in record time—maybe as quickly as a fly can blink its eye.

By B. Toshio Nakatsubo, Senior Managing Director & Executive Director, Corporate Research, Olympus Optical Co., Ltd.

TECHNOLOGIES FOR BETTER HEALTH CARE

Since its inception as the producer of Japan's first microscope, Olympus Optical Co. has focused its business operations on optical technologies, such as those for cameras, endoscopes and analysis equipment. These products, all of which were previously the "stand-alone" type, have gradually come to feature system functions. Images produced by a still video system, for example, are not merely displayed on monitors but are transformed into pictures by printers, and digitally transmitted to another location via public communication lines, or are transferred to and stored on a separate memory unit. The company is currently placing emphasis on developing such electronic and information-processing systems as well as software for operating those systems.

Over the long term, we will seek to respond to the need for multimedia systems. Specifically, the firm will make greater efforts to develop optoelectronic systems, using telecommunication lines. For instance, we plan to create teleconferencing systems that will enable pathologists to make diagnoses by examining microscopic pathological images received from remote locations via telecommunication lines. We hope to produce systems under which pathologists can operate a remote-controlled microscope to conduct studies at a separate location and monitor the work on their own microscopes. Considering the current dearth of pathologists, such systems are expected to greatly contribute to medical services if they can shorten the time required for pathological diagnoses. We also aim to develop an endoscope system integrated with voice, data and character processing units.

Furthermore, Olympus Optical will step up its efforts to develop multimedia equipment and systems for commercial purposes, and for use in specific business areas that the company has cultivated. It is now gearing up its research and development activities for image input and processing technologies, including bandwidth compression.

While focusing our business on optical technologies, we also intend to promote our identity as a manufacturer of precision machinery and equipment. We will direct our energy to research on microtechnologies, such as microoptics and microelectronics, not to mention the so-called micromachines or highly miniaturized electronic devices, which are expected to find a wide range of medical and industrial applications. We hope these efforts will lead to the development of an optical sensor to replace charge coupled devices (CCD), a videoscope that would be combined with a charge modulation device (CMD), and eventually to the development of probe and X-ray microscopes.

We intend to apply our optical technologies not only to develop medical diagnosis equipment but also to design devices for use in preventive medical services. We are now developing a system to store medical examination data on optical cards as well as a dietary consultation system that records on optical cards the history of a patient's eating habits and thus helps dietitians provide advice on health.

In promoting research and development in these fields, we have been strengthening our cooperation with affiliated overseas institutions. Our Biomedical Research Center in Tokyo and an affiliate in New York are undertaking joint studies on reagent and examination methods.

We also promote the exchange of technologies and joint studies in new areas with foreign companies. We have already established a joint venture with Symbol Technologies Inc. in New York to tackle the development of two-dimensional bar codes. In the medical care field, we have formed a business link with Ethicon Emdo Surgery Inc., a New Jersey subsidiary of the Johnson & Johnson group, to manufacture and market instruments for surgical operations.

OUR OPTICAL TECHNOLOGIES WILL BE USED TO DEVELOP STATE-OF-THE-ART MEDICAL DIAGNOSIS EQUIPMENT.

Furthermore, Olympus Optical regularly hires non-Japanese researchers and accepts trainees from the United States and Europe. Introduction of personnel from overseas has proved beneficial to our company in promoting our global business operations.

In the field of basic research, the U.S. is running ahead of Japan. In order to learn from the U.S., we annually send two or three researchers to America for study.

It is our earnest hope that such international cooperation will enable our company to promote a fusion of optical technologies, microtechnologies and information-related technologies, and to further contribute to better the health of mankind.

No limits.

There's very little you can't do with the Olympus Zoom Lens Reflex series. The all-in-one cameras that won't limit your creative ability.

Let's start with the fact that each camera is fully automatic. If you want. If not, you have complete control thanks to manual override features.

Next, you'll find shutter speeds of 1/2000 sec. A versatile dual-element Intelligent Flash System. Special modes for special situations. Plus an LCD viewfinder panel that keeps you informed.

Finally, with the IS-1000 and IS-2000 you get a built-in 35mm – 135mm zoom lens, and on the IS-3000 it's a 35mm – 180mm zoom. All made with our exclusive ED (Extraordinary Dispersion) glass that lets you take telephoto and macro close-ups that are both clear and sharp.

The Olympus IS-1000, IS-2000 and IS-3000. There's no limit to what you can do.

iS-1000

iS-2000

iS-3000

OLYMPUS®

OLYMPUS OPTICAL CO., LTD. Tokyo, New York, Hamburg, London

United Kingdom: OLYMPUS OPTICAL CO. (U.K.) LTD.
2-8 Honduras Street, London EC1Y 0TX Tel: (071) 2532772

Learning from Asian Schools

American schools could benefit from the teaching styles and institutional structures used in Asia—many of which were pioneered here

by Harold W. Stevenson

During the past decade, it has become a truism that American students are not being adequately prepared to compete in a global economy. The latest research shows that the deficiencies become apparent as early as kindergarten and persist throughout the school years. These deficiencies have been most evident when the students are compared with their peers in East Asia. Yet contrary to popular stereotypes the high levels of achievement in Asian schools are not the result of rote learning and repeated drilling by overburdened, tense youngsters. Children are motivated to learn; teaching is innovative and interesting. Knowledge is not forced on children; instead the students are led to construct their own ways of representing this knowledge. The long school days in Asia are broken up by extensive amounts of recess. The recess in turn fosters a positive attitude toward academics.

My colleagues and I gained these insights in a series of five collaborative, large cross-national studies begun in 1980. We explored the children's experiences both at home and at school in the U.S., China, Taiwan and Japan. We found that there is nothing mysterious about the teaching styles and techniques used in Asian schools. Rather these societies embody many of the ideals Americans have for their own schools. They just happen to apply them in an interesting, productive way that makes learning enjoyable.

The vast cultural differences preclude direct translation of many of the practices and beliefs from those cultures to our own. But these comparative data have helped us realize how far Americans have strayed from the effective application of well-known teaching methods. The studies have revealed new perspectives about our own culture and fresh ideas about how our educational system might be improved. Indeed, simply increasing the length of school days would be meaningless if there were no change in the way American teachers are asked to perform their jobs.

HAROLD W. STEVENSON is professor of psychology at the University of Michigan, Ann Arbor. He received his Ph.D. from Stanford University. Since 1979, he has been conducting a series of cross-national studies of children's academic achievement. Stevenson has earned several distinctions, including a Guggenheim fellowship and the American Psychological Association's G. Stanley Hall award for research in developmental psychology.

Results from cross-national studies can be greatly distorted if the research procedures are not comparable in each area and if the test materials are not culturally appropriate. We avoided the first potential problem by selecting a full range of schools in five metropolitan areas: Minneapolis, Chicago, Sendai, Beijing and Taipei. These cities are similar in size and cultural status within their own countries. In each metropolitan region, we selected from 10 to 20 elementary schools that represented a range of students from different socioeconomic backgrounds. (Because socioeconomic status is not easy to define, we used the parents' educational level as the basis for selection.) We then randomly chose two first-grade and two fifth-grade classrooms in each school.

To avoid the difficulty in translating materials developed in one culture for use in another, we constructed our own tests. We began by compiling computer files of every concept and skill included in the students' mathematics textbooks and of every word and grammatical structure in their reading material. With these files, we were able to create test items that were relevant to each culture and that were at the appropriate levels of difficulty.

Armed with these materials, we administered mathematics and reading tests to thousands of students in the first- and fifth-grade classrooms. Later we randomly selected samples of six boys and six girls from each classroom for more in-depth testing and interviews. In one of the studies, we visited

GRADE SCHOOLS in Asia intersperse studying with frequent periods of activity. Such an approach helps to maintain children's attention and may make learning easier and more enjoyable. Here, fourth graders from Japan are making wearable masks.

a total of 204 classrooms in 11 schools in Beijing, 10 in Taipei, 10 in Sendai and 20 in Chicago.

The test results confirmed what has become common knowledge: schoolchildren in Asia perform better academically than do those in the U.S. In mathematics the average scores of the Asian first graders were above the American average, but scores of some of the American schools were as high as some of those in Sendai and Taipei. By fifth grade, however, the American students lost much ground: the average score of only one of the Chicago-area schools was as high as the worst of the Asian schools. On a computation test, for example, only 2.2 percent of the Beijing first graders and 1.4 percent of the fifth graders scored as low as the mean for their Chicago counterparts. On a test of word problems, only 2.6 percent of the Beijing first graders and 10 percent of the fifth graders scored at or below corresponding American means.

The deficiencies of American children appear to build throughout the school years. When we compared the scores of kindergarten children and of first, fifth and 11th graders in Minneapolis, Sendai and Taipei, we found a relative decline in the scores among the American students, improvement in Taiwan and steady high performance in Japan.

American students' shortcomings are not limited to mathematics. Although Americans performed the best on reading in the first grade, the Asian students had caught up by the fifth grade. The rise is remarkable when one considers the reading demands of Asian languages. Chinese students had to learn several thousand characters by the fifth grade, and Japanese students had to learn Chinese characters, two syllabaries (symbols for the syllables in Japanese) and the roman alphabet.

Because of the early onset and pervasiveness of cross-cultural differences in academic achievement, it seemed obvious that we would have to investigate attitudes, beliefs and practices related to children's success. We spent hundreds of hours observing in the classrooms, interviewed the teachers, children and mothers and gave questionnaires to the fathers.

American parents show a surprisingly high level of satisfaction with their children's level of academic performance. From kindergarten through the 11th grade, more than three times as many Minneapolis mothers as Asian mothers said they were very satisfied with their child's current level of achievement.

The U.S. students were also very positive about their abilities. More than 30 percent of the Chicago fifth graders considered themselves to be "among the best" in mathematics, in reading, in sports and in getting along with other children. Such self-ratings were significantly higher than those made by Sendai

and Taipei children for mathematics and by Sendai children for reading. Taipei children gave the highest self-ratings for reading. Except for social skills, many fewer Beijing children gave themselves such positive ratings.

In another set of questions, we asked the mothers how well the school was educating their own children. More than 80 percent of the American mothers expressed a high level of satisfaction. Except at kindergarten, when mothers in all four societies were quite satisfied, Minneapolis mothers felt much better about their children's schools than did mothers in Taipei and Sendai.

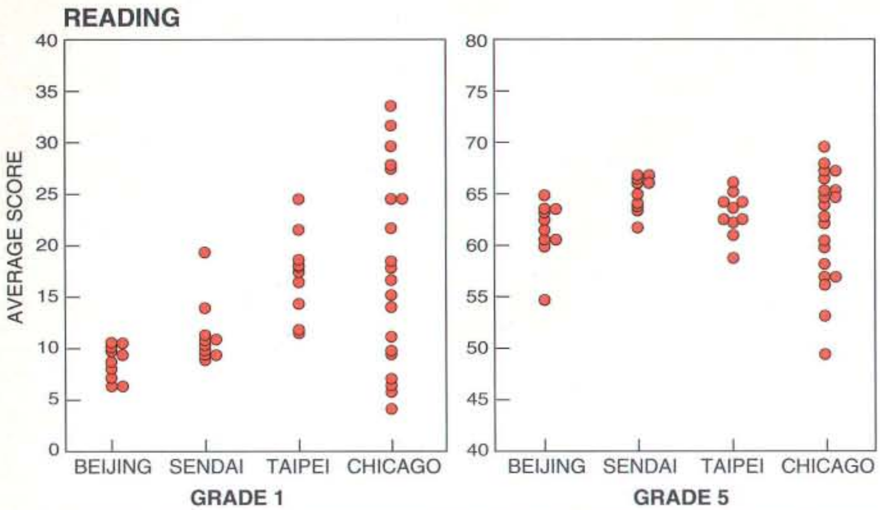
Why should American mothers be so positive? One likely explanation is that they lack clear standards to which they can refer. No national or state curricula define what children should learn at each grade, and few mothers receive more than vague reports about their children's performance. American mothers also seem to place a lesser emphasis on academic achievement. In the U.S., childhood is a time for many different types of accomplishment. Doing well in school is only one of them.

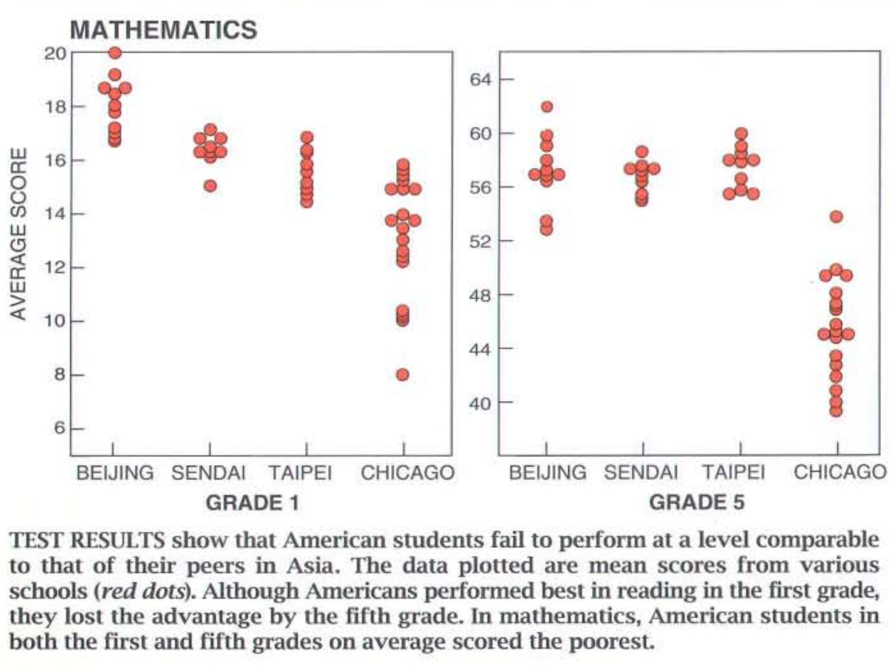
Asian mothers, on the other hand, have told us repeatedly that their children's primary task is to do well in school. The mothers' own job is to try to do everything possible to ensure that success. They regarded education as critical for their children's future. Thus, Asian mothers find it more difficult to be satisfied with moderate levels of performance.

The American mothers' contentment had clearly been transmitted to their children. Fifth graders were asked whether they agreed with the statement, "I am doing as well in school as my parents want me to." American children thought this statement was more true of them than did the Asian children. We obtained similar results when we asked the question in terms of their teachers' satisfaction.

The American mothers we interviewed apparently were not strongly impressed by recent criticisms of U.S. education. As far as they were concerned, the relatively poor academic showing of U.S. students did not reflect the abilities of their own children or their children's schools. For American mothers, problems existed at other schools and with other children. Our interviews revealed little evidence that American mothers were motivated to seek improvements in the quality of their children's education or that American children believed they were doing anything but a satisfactory job in school.

We explored academic motivation in another way by posing a hypothetical question to the children: "Let's say there is a wizard who will let you make a wish about anything you want. What would you wish?" The most frequent wishes fell into four categories: money; material objects, such as toys or pets; fantasy, such as wanting to be sent to the moon or to have more wishes; and educational aspirations, such as doing well in school or going to college. Almost 70 percent of

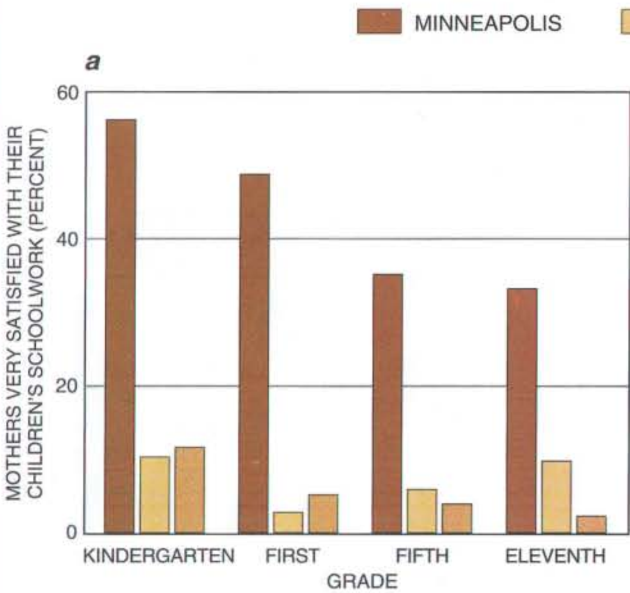


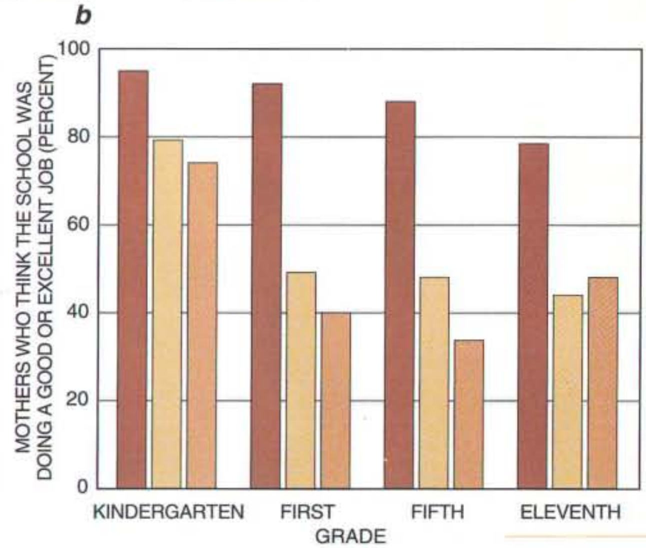


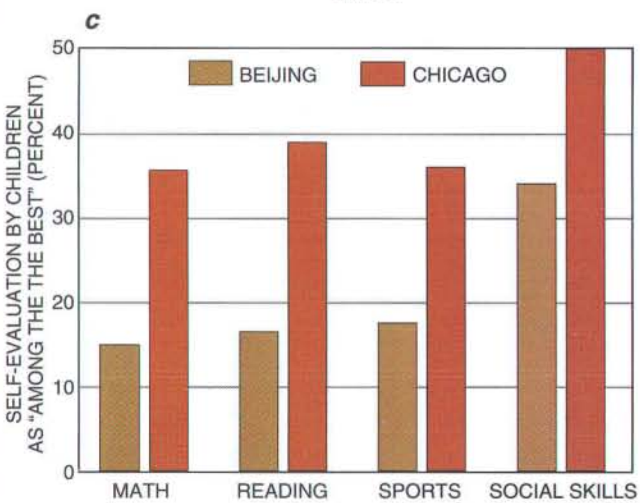
TEST RESULTS show that American students fail to perform at a level comparable to that of their peers in Asia. The data plotted are mean scores from various schools (*red dots*). Although Americans performed best in reading in the first grade, they lost the advantage by the fifth grade. In mathematics, American students in both the first and fifth grades on average scored the poorest.

How Parents and Students See the System

Satisfaction with academic performance of students (*a*) and their children's schools (*b*) is higher among American mothers that it is among Asian mothers. Children seem to reflect their parents' attitudes: more American students rated themselves "among the best" in several categories than did children from Beijing (*c*).







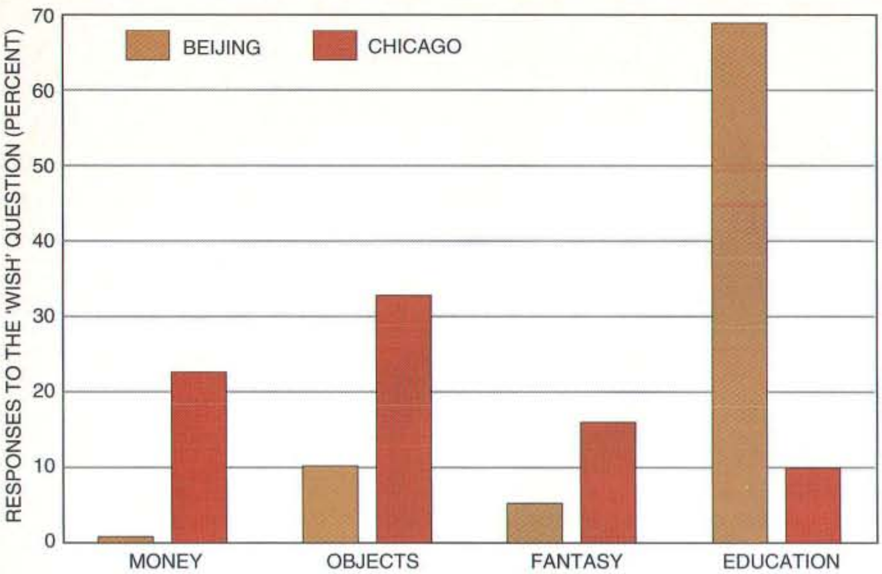
the Chinese children focused their wishes on education. American children were more interested in receiving money and material objects. Fewer than 10 percent of the American children expressed wishes about education.

The enthusiasm Asian children express about school comes in part, of course, from the well-known societal emphasis on education. Several studies of immigrants have documented the willingness of Asian children to work hard [see "Indochinese Refugee Families and Academic Achievement," by Nathan Caplan, Marcella H. Choy and John K. Whitmore; SCIENTIFIC AMERICAN, February]. This attitude stems from Confucian beliefs about the role of effort and ability in achievement. The malleability of human behavior has long been emphasized in Chinese writings, and a similar theme is found in Japanese philosophy. Individual differences in potential are deemphasized, and great importance is placed on the role of effort and diligence in modifying the course of human development.

In contrast, Americans are much more likely to point to the limitations imposed by an assumed level of innate ability. This belief has potentially devastating effects. When parents believe success in school depends for the most part on ability rather than effort, they are less likely to foster participation in activities related to academic achievement. Such parents may question whether spending time in academic pursuits after school is useful for children of presumed low ability. They may readily accept poor performance. Furthermore, if the parents believe the child has high ability, they may question whether such activities are needed.

It is relatively easy to demonstrate the greater emphasis placed by Americans on innate ability. One approach is to ask children to rate the importance of certain factors for doing well in school. Beijing children emphasized effort rather than ability. Chicago children thought both to be of near-equal importance.

In another approach we asked Taipei, Sendai and Minneapolis children to in-



WISHFUL THINKING seems to reflect cultural priorities. Chicago children tended to wish for money and material objects, such as toys or pets. Beijing children wished for educational goals, such as going to college.

dicate the degree to which they agreed with the statement that "everybody in the class has about the same amount of ability in math." American children expressed less strong agreement than did the Chinese and Japanese children. Mothers follow the same pattern of response. When asked about the degree to which they agreed that "any student can be good at math if he/she works hard enough," Minneapolis mothers expressed less agreement than did Sendai or Taipei mothers.

Children may work harder because they believe achievement depends on diligence. The idea that increased effort will lead to improved performance is an important factor in accounting for the willingness of Chinese and Japanese children, teachers and parents to spend so much time and effort on the children's academic work.

The enthusiasm for school also seems to come from the happy times that Asian children appear to have there. That these children regard school as pleasant—rather than as regimented, austere and demanding—surprises most Americans. Our stereotype is that the intense quest for academic excellence reduces the possibilities of making school a place that children enjoy. This clearly is not the case.

When we compare the daily routine in Asian and American schools, we realize how easy it is to overlook the constraints of American schools. Classes begin shortly after the students arrive, and the children leave just after their last class ends. Rarely is there more than a single recess. The lunch period—a potential time for play and social interaction—is usually limited to half an hour or less. As a consequence, American children spend most of their time at school in the classroom.

In contrast, the daily routine in Asian schools offers many opportunities for social experience. There are frequent recesses, long lunch periods and after-school activities and clubs. Such opportunities make up about one fourth of the time spent during the eight hours at school. The school day is longer in Asia mainly because of the time devoted to these nonacademic periods. Play, social interaction and extracurricular activity may not contribute directly to academic success, but they make school an enjoyable place. The enjoyment likely creates cooperative attitudes.

The relative lack of nonacademic activities in American schools is reflected in greater amounts of time spent in play after school. American mothers estimated that their elementary school children spend nearly 25 hours a week playing. We found this surprisingly high until we considered how little time was available for play at school. Estimates for Chinese children were much lower—their social life at school is reflected in the shorter amounts of time they play after returning home.

Chicago children also spent nearly twice as much time as Beijing children watching television. But compared with Americans, Japanese students from kindergarten through high school spent even more time watching television. The difference in this case appears to be that Japanese children are more likely to watch television after they had completed their homework.

American children were reported to spend significantly less time than Asian children in doing homework and reading for pleasure—two pursuits that are likely to contribute to academic achievement. Mothers estimated that the Taipei children spent about four times as much time each day doing homework as did American children and over twice as much time as did Japanese children. American children were estimated to spend less time reading for pleasure than their Asian peers throughout their school years.

The enjoyment Asian students have at school may be the reason they appear to Western visitors as well behaved and well adjusted. These observations, however, have always been informal, and no data exist to support or refute them. So we decided to ask mothers and teachers in Beijing and Chicago about these matters. In particular, we questioned them about physical symptoms of tension, which we thought would be a good indicator of adjustment.

Chinese mothers reported fewer complaints by their children of stomachaches and headaches, as well as fewer requests to stay home from school than did American mothers. The Chinese mothers also more frequently described their children as happy and obedient. Only 4 percent of the Chinese mothers, but 20 percent of the American mothers, said their children encountered problems in getting along with other children.

The intense dedication of Chinese elementary school children to schoolwork did not appear to result in tension and maladjustment. Nor have we found patterns of psychological disturbance among several thousand Chinese and Japanese 11th graders in self-evaluations of stress, depression, academic anxiety or psychosomatic complaints. Our data do not support the Western assumption that Asian children must experience extraordinary stress from their more demanding curriculum. The clear academic goals and the enthusiastic support given by family, teachers and peers may reduce the strain from working so hard.

The achievement of Asian students is facilitated by the extensive amount of attention teachers can give the children. Indeed, one of the biggest differences we found was the amount of time teachers had. Beijing teachers were incredulous after we described a typical day in American schools. When, they asked, did the teachers prepare their lessons, consult with one another about teaching techniques, grade the students' papers and work with individual students who were having difficulties? Beijing teachers, they ex-

plained, are responsible for classes for no more than three hours a day; for those with homeroom duties, the total is four hours. The situation is similar in Japan and Taiwan, where, according to our estimates, teachers are in charge of classes only 60 percent of the time they are at school.

Teaching is more of a group endeavor in Asia than it is in the U.S. Teachers frequently consult with one another, because, in following the national curriculum, they are all teaching the same lesson at about the same time. More experienced teachers help newer ones. Head teachers in each grade organize meetings to discuss technique and to devise lesson plans and handouts. The group may spend hours designing a single lesson or discussing how to frame questions that will produce the greatest understanding from their pupils. They also have a teachers' room, where all the instructors have desks and where they keep their books and teaching materials. They spend most of the time there when not teaching.

American teachers have neither the time nor the incentive to share experiences with one another or to benefit from hearing about the successes and failures of other instructors. Each teacher's desk is in the classroom, and little space is allocated specifically for informal discussions and meetings. The teachers' room in American schools is typically a place to rest rather than to work. As a result, American teachers spend most of their time at school isolated in their own classrooms, with few opportunities for professional interaction or consultation.

With no national curriculum or guidelines, American schools typically develop their own agenda. In any year the curriculum may not be consistent within a city or even within a single school. Adding further to the diversity in the curricula among American classrooms is the fact that teachers are free to proceed through textbooks at any rate they wish, skipping the parts they do not find especially interesting or useful.

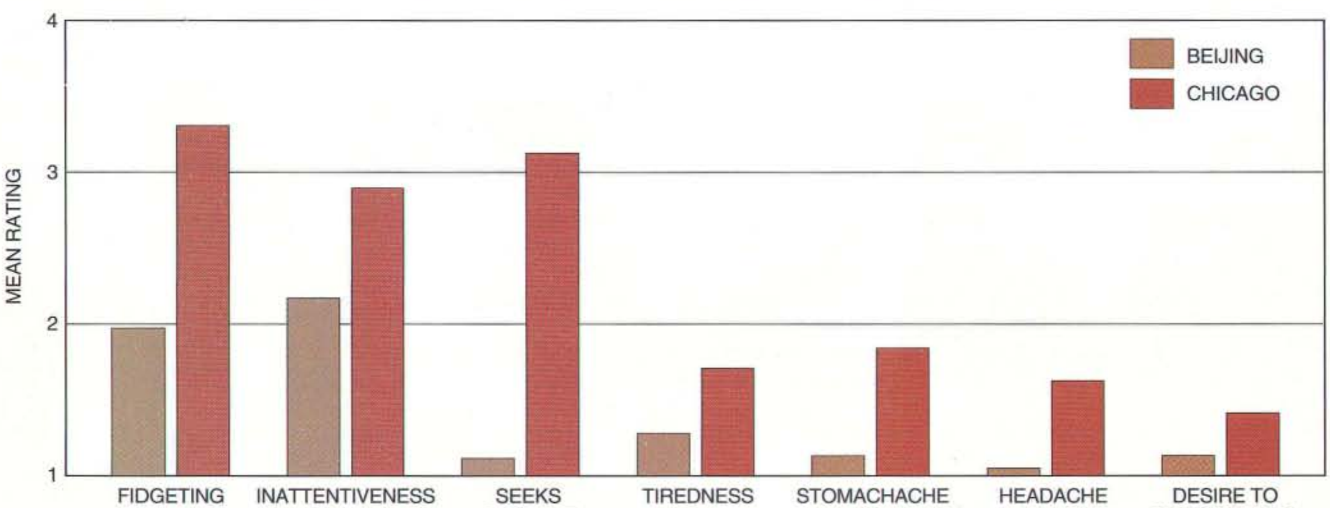
The demanding daily schedule places serious constraints on the ability of American teachers to create exciting, well-organized lessons. They usually must prepare for their classes at home during evenings and weekends. Furthermore, they must cover all elementary school subjects, because teachers for specific academic subjects typically do not appear until junior high school. Evenings are not the most appropriate time to begin such a difficult task, for teachers are tired from the demands of school and their own affairs. There are, of course, excellent American teachers. And there are individual differences among Asian teachers. But what has impressed us in our observations and in the data from our studies is how remarkably well most East Asian teachers do their jobs.

Asian teachers can be described best as well-informed, well-prepared guides. They do not see themselves primarily as dispensers of information and arbiters of what is correct but rather as persons responsible for guiding students skillfully through the material. Each lesson typically begins with a statement of its purpose and ends with a summary of its content. The lesson follows a well-planned script in which children are led through a series of productive interactions and discussions. Teachers regard children as active participants in the learning process who must play an important role in producing, explaining and evaluating solutions to problems.

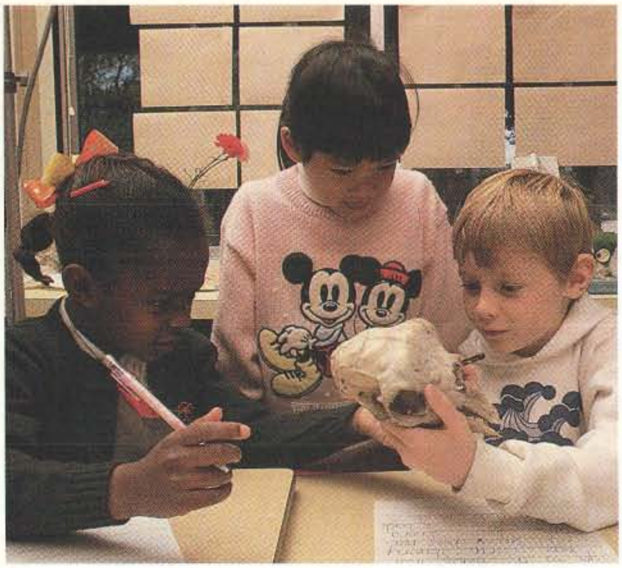
The skill shown by Asian teachers is not acquired in college. In fact, some teachers in China have only a high school education. The pattern for training teachers resembles that provided to other professionals: in-service training under the supervision of skilled models. Colleges are assumed to provide basic knowledge about subject matter, as well as about child development and theories of learning. But Asian instructors believe the art of teaching can be accomplished better in classrooms of elementary schools than in lecture halls of colleges. This approach stands in sharp contrast to that taken in the U.S., where teaching skill is generally thought to be best acquired through several specialized courses in teaching methods.

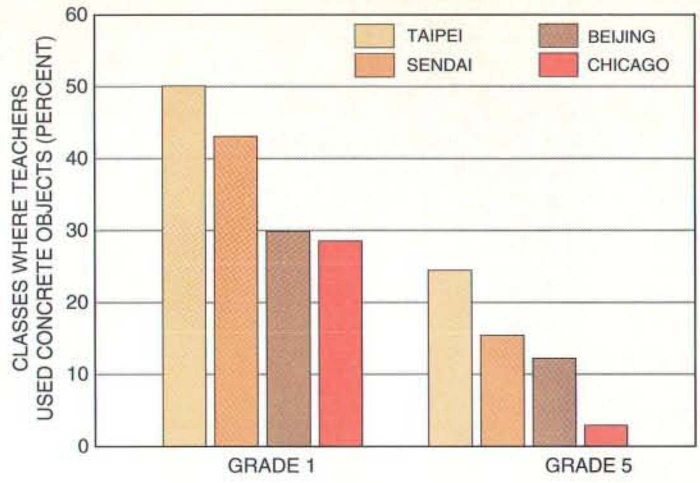
The skills employed by Asian teachers are also more effective in attracting and maintaining children's attention. We found Asian children listening to the teacher more frequently than American children—at least 80 percent of the time versus approximately 60 percent. This finding may also result from differences in the number of recesses in Asian and American schools. Attention is more likely to falter after several hours of classes than it is if opportunities for play and relaxation precede each class, as is the case in Asian schools.

Another likely reason for the American children's lack of attention lies in the manner in which U.S. teachers often structure the lessons. Because of the time spent on seatwork—exercises or assignments children are to complete at their desks—and the way in which seatwork is used, American children have fewer opportunities to interact with the teacher than do children in Asian



SYMPTOMS OF STRESS in U.S. classrooms exceed those in Asian schools. First-grade teachers in Chicago and Beijing estimated the frequency of physical complaints among students. They rated their impressions on a scale from 1 (seldom complained) to 5 (often complained). American schoolchildren expressed symptoms more often than did Beijing students.





USING OBJECTS to teach young children is often more effective than verbal instruction. In mathematics classes, teachers in Asia use concrete materials more frequently than do American teachers.

classrooms. American teachers rely more heavily on seatwork than do Asian teachers, which is not surprising. Giving children such tasks is one of the few ways American teachers can gain some free time in their grueling daily schedule.

In the U.S., teachers usually explain the concepts during the early part of the lesson and then assign seatwork during the later part. Asian teachers, on the other hand, intersperse brief periods of seatwork throughout the class period. Seatwork is a means of getting children to practice what they have just learned and of quickly spotting difficulties the children might be encountering. American teachers are less likely to take advantage of the diagnostic value of seatwork.

Children's attention also increases when they receive feedback. If students do not receive some type of acknowledgment from the teacher or some indication of whether their work is correct, they are more likely to lose interest. In a surprisingly large number of classes, Chicago teachers failed to provide feedback to the children, especially when the children were doing seatwork. In nearly half of the 160 class periods we observed in the Chicago fifth grades, teachers failed to offer any type of evaluation as the children worked alone at their desks. In striking contrast, lack of such acknowledgment was practically never observed in Sendai and only infrequently in Taipei.

In addition, Asian teachers make more frequent use of materials that can be manipulated. Jean Piaget and other psychologists (as well as most parents and teachers) have discovered that young children enjoy manipulating concrete objects, which is often a more effective way to learn than is listening to verbal instruction. Even so, American teachers were much less likely than Asian teachers to provide concrete objects for manipulation in mathematics classes.

Finally, Asian teachers make the subjects interesting by giving them some meaningful relation to the children's everyday lives. In mathematics, word problems often serve this function. They can transform mathematics from a subject of dull computation to one requiring active problem solving. We found that American teachers were much less likely than Asian teachers to introduce word problems. Fifth-grade American teachers presented word problems in less than one out of five class periods; Sendai teachers included word problems in more than eight out of every 10 lessons. Similar differences emerged when we calculated how often children were asked to construct word problems themselves. This exercise rarely occurred in American classrooms.

Asian teachers are able to engage children's interest not because they have insights that are unknown in the U.S. but because they take well-known principles and have the time and energy to apply them with remarkable skill. They incorporate a variety of teaching techniques within each lesson, rely more frequently on discussion than on lectures, teach children how to make smooth transitions from one type of activity to another and seldom engage in irrelevant discussions—all approaches to teaching that American instructors would agree are reasonable and effective.

Perhaps the most pointed difference between the goals of Asian and American teachers emerged when we asked teachers in Beijing and Chicago what they considered to be the most important characteristics of a good instructor. "Clarity," said nearly half of the Beijing teachers. "Sensitivity to the needs of individuals" was the most common response of the Chicago teachers. Beijing teachers were also more likely to emphasize enthusiasm, and Chicago teachers were more likely to stress patience.

Have the goals of education diverged to such a degree in Eastern and Western cultures that American teachers see their main tasks as those of evaluating and meeting the needs of individuals, while Asian teachers can devote their attention to the process of teaching? If this is the case, the academic achievement of American children will not improve until conditions are as favorable as those provided in Asia. Clearly, a challenge in the U.S. is to create a greater cultural emphasis on education and academic success. But we must also make changes in the training of teachers and in their teaching schedules, so that they, too, will be able to incorporate sound teaching practices into their daily routines.

FURTHER READING

PRESCHOOL IN THREE CULTURES: JAPAN, CHINA, AND THE UNITED STATES. J. J. Tobin, D. Y. H. Wu and D. H. Davidson. Yale University Press, 1989.

JAPANESE EDUCATIONAL PRODUCTIVITY. Edited by R. Leestma and H. J. Walberg. University of Michigan, Center for Japanese Studies, 1992.

LEARNING TO GO TO SCHOOL IN JAPAN: THE TRANSITION FROM HOME TO PRESCHOOL LIFE. L. Peak. University of California Press, 1992.

THE LEARNING GAP: WHY OUR SCHOOLS ARE FAILING AND WHAT WE CAN LEARN FROM JAPANESE AND CHINESE EDUCATION. H. W. Stevenson and J. W. Stigler. Summit Books, 1992.

Extremely Cold Antiprotons

Cooling and trapping of these particles at energies one ten-billionth of what was feasible six years ago should make possible production of the first antimatter atoms

by Gerald Gabrielse

How does one store an antiproton? The antimatter counterpart of the familiar proton (a building block of ordinary matter, along with the electron and the neutron), the antiproton is believed to have the same amount of charge as the proton, but the charge is negative instead of positive. A single collision between a proton and an antiproton can annihilate both particles. In a burst of energy the antiproton and proton cease to be, and a variety of particles (most of which are called pions) are formed. As a consequence, antiprotons cannot be confined by the walls of an ordinary container, nor can they come into contact with the ordinary atoms making up the atmosphere. The only way to store an antiproton is in a nearly perfect vacuum, using magnetic and electric fields to make a container without walls.

The past few years have seen a remarkable blossoming in the ability to cool and store antiprotons. Storage techniques involving magnetic fields have been common since 1955, when the Bevatron storage ring at the University of California at Berkeley was constructed to confine antiprotons. Yet the antiprotons in such large rings are extremely "hot": they typically travel at speeds approaching the limiting speed of light and have extremely high energies, ranging from a billion to a trillion electron volts (1 GeV to 1 TeV). High speeds and energies are acceptable and even desirable for experiments in which antiprotons collide with other particles, but other interesting experiments require "cold" antiprotons that move slowly and have low energies. Such experiments are needed in order to test accurately our understanding of matter and antimatter, as well as theories that underlie the nature of both substances.

Recently a small international team, of which I am a member, demonstrated the ability to slow and cool antiprotons to energies one ten-billionth of what was possible just six years ago. The cold antiprotons can be stored for as long as desired, even for several months, in a nearly perfect vacuum that is less than a cubic millimeter in volume. The average energy of the antiprotons is so low—less than one thousandth of an electron volt—that it is typically expressed in terms of temperature units. (A fortieth of an electron volt corresponds to room temperature.) The antiprotons in our storage device have a temperature of only four degrees above absolute zero (four kelvins).

Extremely cold antiprotons are already being exploited to compare the charges and masses of antiprotons and protons at a level of accuracy more than 1,000 times greater than was previously possible. Such comparisons stringently test the so-called *PCT* theorem of particle physics, which predicts that the antiproton and proton should have identical masses and charges that differ only in sign. We expect a substantial improvement in accuracy over the next several years. Someday cold antiprotons might even be used to observe the first antimatter atoms. By combining an antiproton with a positron (an antielectron), it should be possible to produce antihydrogen.

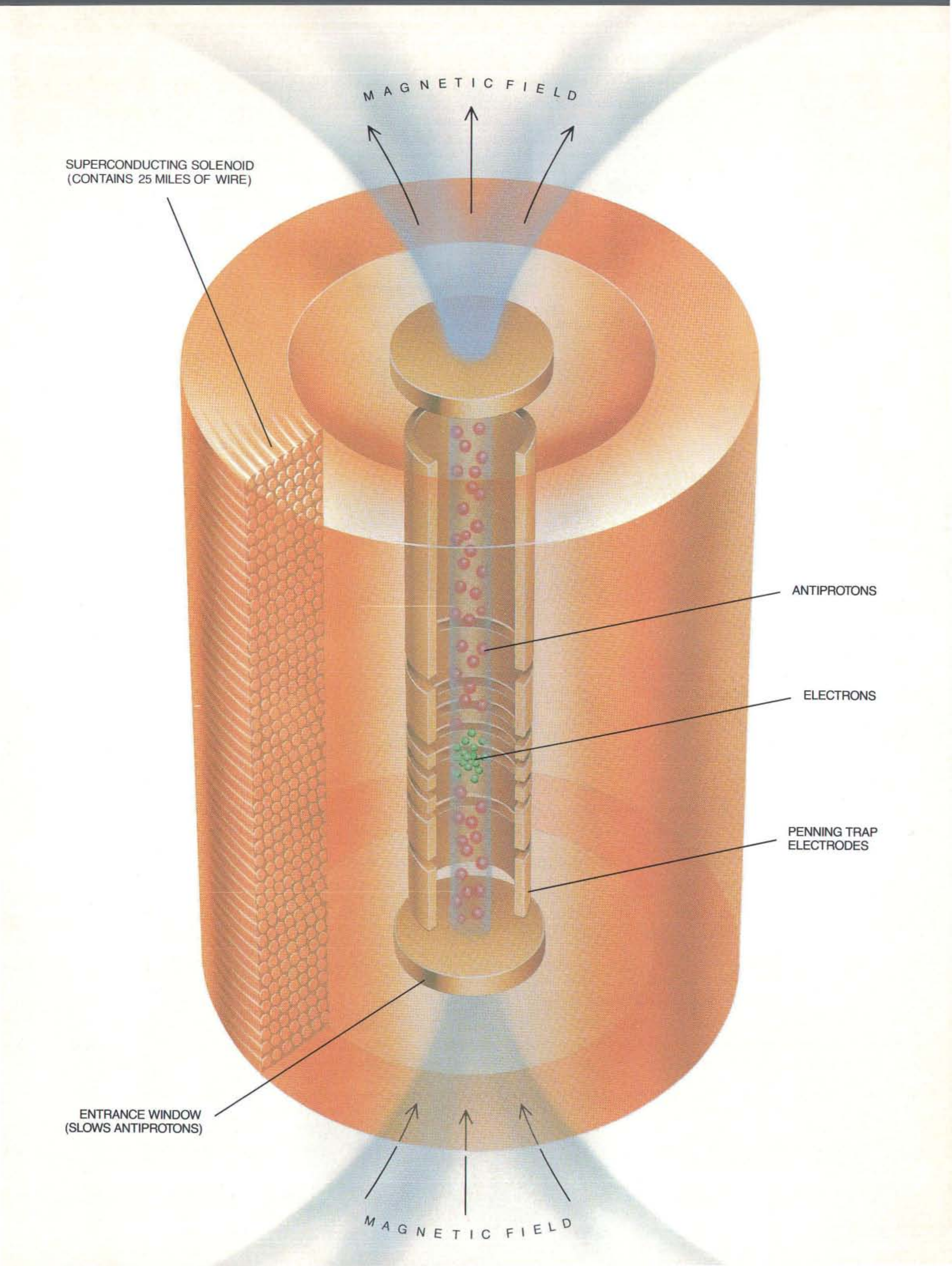
Antiprotons occur naturally only as the rare products of collisions between high-energy cosmic rays and atoms in the atmosphere. Although they are believed to be stable—that is, they do not spontaneously decay into other particles—such naturally occurring antiprotons nonetheless live for only a very short time. Soon after they come into being, they annihilate in collisions with protons that are in the atmosphere.

Antiprotons are created artificially in particle accelerators by colliding extremely high energy protons with solid matter. CERN, the European laboratory for particle physics near Geneva, generated collisions between large numbers of antiprotons and protons to observe and study the short-lived *W* and *Z* particles. At the Fermi National Accelerator Laboratory in Batavia, Ill., higher-energy collisions between antiprotons and protons are now being investigated as a continuation of a long search for the top quark. This particle is the only member of a group of six constituents of heavy particles (such as protons) that has not been observed.

New experiments with antiprotons became possible when workers at CERN scavenged parts from earlier storage rings to complete the Low Energy Antiproton Ring (LEAR) in 1982. LEAR has a modest circumference of only 79 meters, which is tiny compared with the 85-kilometer circumference of the contemplated 20-TeV Superconducting

GERALD GABRIELSE is professor of physics at Harvard University, where he currently enjoys teaching a graduate course in quantum mechanics. He received a B.S. degree from Calvin College in 1973 and a Ph.D. from the University of Chicago in 1978 and then was on the faculty of the University of Washington until 1987. He has led the experimental efforts to produce and use extremely cold antiprotons from their inception. He investigates single electrons and coupled systems of electron oscillators, along with performing mass spectroscopy, positron experiments and recombination studies related to the antiproton experiments. He is husband to Ellen and father to Abigail, Joshua and Deborah, with whom he likes to backpack in the mountains.

ION TRAP captures antiprotons (*red*), which are cooled by collisions with cold electrons (*green*). The antiprotons and electrons are held by electric and magnetic fields; the electric field is produced by applying voltages to electrodes, and the magnetic field is generated by a superconducting solenoid, or coil (not to scale).



MAGNETIC FIELD
SUPERCONDUCTING SOLENOID
(CONTAINS 25 MILES OF WIRE)
ANTIPROTONS
ELECTRONS
PENNING TRAP
ELECTRODES
ENTRANCE WINDOW
(SLOWS ANTIPROTONS)
MAGNETIC FIELD

Super Collider. LEAR regularly slows and cools antiprotons to an energy of six million electron volts (MeV), which corresponds to a speed that is approximately 10 percent that of light.

Our apparatus is able to cool antiprotons to energies one ten-billionth of those obtained at LEAR, and it is so small that it falls in the realm of the "tabletop." A major complication is that the tabletop must be connected to a series of large particle physics accelerators, such as the machines at CERN, that are capable of supplying antiprotons having energies of a few MeV.

In 1981 I visited Fermilab to explore the possibility of trapping and storing extremely cold antiprotons. Fermilab was closer to home than was CERN, and it had a small storage ring in operation, which seemed adaptable to the project. Unfortunately, the intense focus on studying high-energy collisions between protons and antiprotons left little room for the low-energy experiments envisioned. Cooler heads did not prevail. By 1984, pieces of the small storage ring were being trucked to other laboratories. William Kells of Fermilab and I thus turned our attention to mounting an experiment in Geneva, since the LEAR facility had now become the only laboratory in the world that could slow antiprotons to the MeV-range energies our proposed tabletop apparatus could accept. Hartmut Kalinowsky of the University of Mainz joined forces with us, as did Thomas A. Trainor of the University of Washington.

Oversight committees and administrators at CERN greeted our proposals with some skepticism. We sought to slow the antiprotons in matter, capture them in an ion trap and cool them through collisions with cold electrons in the same trap. (An ion trap confines charged particles, or ions, by means of magnetic and electric fields.) These unproved techniques were quite different from the usual high-energy collision experiments. Moreover, one of our physics goals was in direct competition with a proposed experiment (a large radio-frequency mass spectrometer) in which CERN had already invested a great deal of time and money. There was also much concern because we had no financial support. At the same time, agencies in the U.S. were cautious about funding a large new program to be done at CERN, one that did not yet have approval there. Fortunately, the Atomic Physics Division of the National Science Foundation, followed by the Air Force Office of Scientific Research and the National Bureau of Standards, decided to fund the request for low-energy antiprotons. Somewhat later the German State Ministry for Research joined in.

In May 1986, CERN granted us 24 hours of access to LEAR antiprotons to demonstrate that it actually was possible to slow them from several million to a few thousand electron volts. Our demonstration worked, and we were rewarded by a second 24-hour access period two months later, in which we sought to show that we could capture the slowed antiprotons.

Unfortunately, we had insufficient time to obtain modern equipment with which to build an ion trap. Borrowing an ancient superconducting magnet, we constructed a trap in one day, relying on glass-to-copper seals of unknown origin that we found abandoned in a glassblower's drawer. Our trap was chilled to a temperature of four kelvins by thermal contact with liquid helium within a dewar, a vacuum-insulated vessel similar to a thermos bottle. After testing the apparatus in the U.S., we shipped it by air to CERN because of the pressure of time and the delicate nature of the dewar we had built. Only after the dewar arrived broken in Geneva did we learn that our "air" shipment had in fact rattled across Europe by truck.

Repairing the abused dewar turned out to be an exercise in improvisation. It is hard to forget aiming a sputtering hand torch at "borrowed" high-temperature solder placed on thin tubes, within an apparatus that dangled from a rope tied to an exposed beam in a CERN hallway. The repaired apparatus was ready several days before the antiprotons were scheduled to arrive, which was at noon on Friday, July 17. Feverish computer programming proceeded, punctuated by calls of "just one half-hour more of BASIC" as we sought to interface our computer with devices at LEAR in order to read out information about attempted antiproton captures as they happened.

Then, late Thursday evening, disaster struck. Routine tests unexpectedly revealed that we could no longer apply high voltages to our ion trap without causing an unwanted electric arc deep inside the coldest part of the apparatus. It was 12 hours before the antiprotons were scheduled to arrive, and this apparatus had never been warmed to room temperature and then cooled back to four kelvins in less than several days. Half of our team gave up and went to bed.

Given CERN's ambivalence about the feasibility of the proposed experiments, a failure would clearly be a major setback. A repair had to be attempted. As we opened the cold apparatus, water that had condensed on the "super"-insulation streamed out, despite the hot air directed on it from three industrial-strength hair dryers. Eventually we eliminated the arc by installing fresh cables to handle the high voltages. After much mopping of water, drying and cleaning, we reassembled the apparatus and began cooling it by 10 A.M. Friday.

We told the LEAR control room that shortly after noon we would indeed be ready for our antiproton test. Our exhausted euphoria was short-lived. We were told by telephone that although antiprotons were available in one of the large storage rings at CERN, the "kicker" used to extract antiprotons from the ring had failed. We would most likely have to leave CERN without receiving antiprotons. The test experiment seemed doomed, since LEAR was shortly scheduled to be shut down for more than a year. I made known the urgency of our situation, then stumbled off to bed.

Several hours later I was awakened. A particle accelerator "magician" at

Energy Units

The electron volt (eV) is the energy acquired by an electron when it travels from the negative to the positive terminal of a one-volt battery. An eV is the typical unit of energy used to describe electrons bound in atoms. Standard metric prefixes are added to represent the larger and smaller energies needed to describe the experiments that are discussed in this article:

1 TeV =	1,000,000,000,000 eV =	10^{12} eV
1 GeV =	1,000,000,000 eV =	10^{9} eV
1 MeV =	1,000,000 eV =	10^{6} eV
1 keV =	1,000 eV =	10^{3} eV
1 meV =	0.001 eV =	10^{-3} eV

Small energies are sometimes represented in temperature units, in degrees above absolute zero (K), with 1 meV ~ 12 K. The much larger GeV and TeV, used to describe the energy of accelerated particles, are still very small compared with the kinetic energy ($E = \frac{1}{2} Mv^2$) of macroscopic objects of mass M and speed v. For example, a one-gram paper clip dropped one meter strikes the ground with a kinetic energy on the order of 10^{17} eV = 10^{5} TeV.

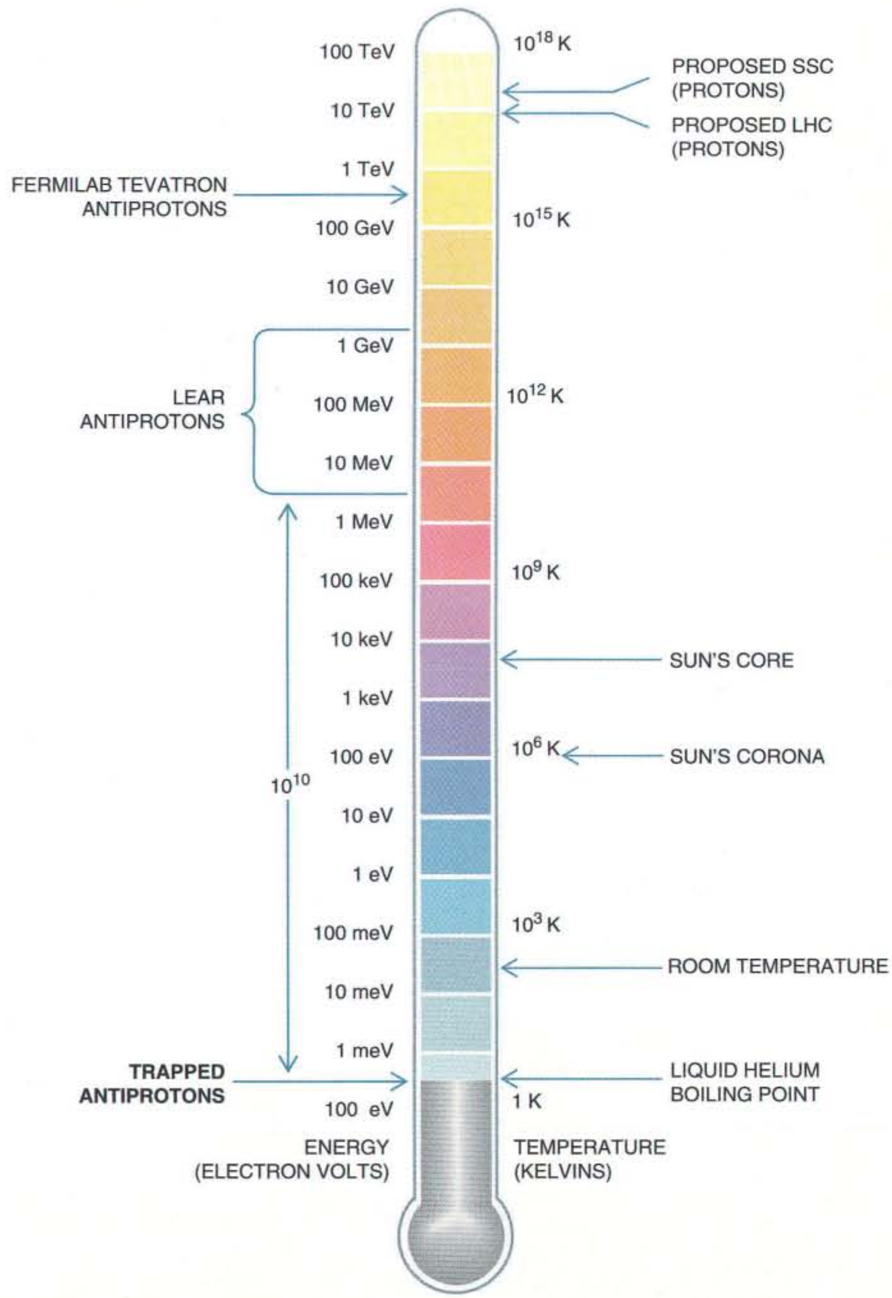
CERN had managed to make a backup "kicker" work for the first time. Soon LEAR was ready to send us brief, intense pulses, or bursts, of antiprotons. (Typically each pulse contained 100 million antiprotons and had a duration of 200 billionths of a second.) Operators counted "five, four, three, two, one" in various versions of English and then pushed a newly installed green button with a loud "go." After several hours of adjusting the timing electronics, we observed pions from the annihilation of antiprotons that we had trapped briefly and then released.

The emotional rollercoaster ended on a pronounced high. LEAR operators and physicists from other experiments crowded around the console during the countdown. Applause broke out whenever the histogram on the computer monitor indicated that antiprotons had been trapped and stored. A few antiprotons were held for 20 minutes, establishing the feasibility of the proposed experiments.

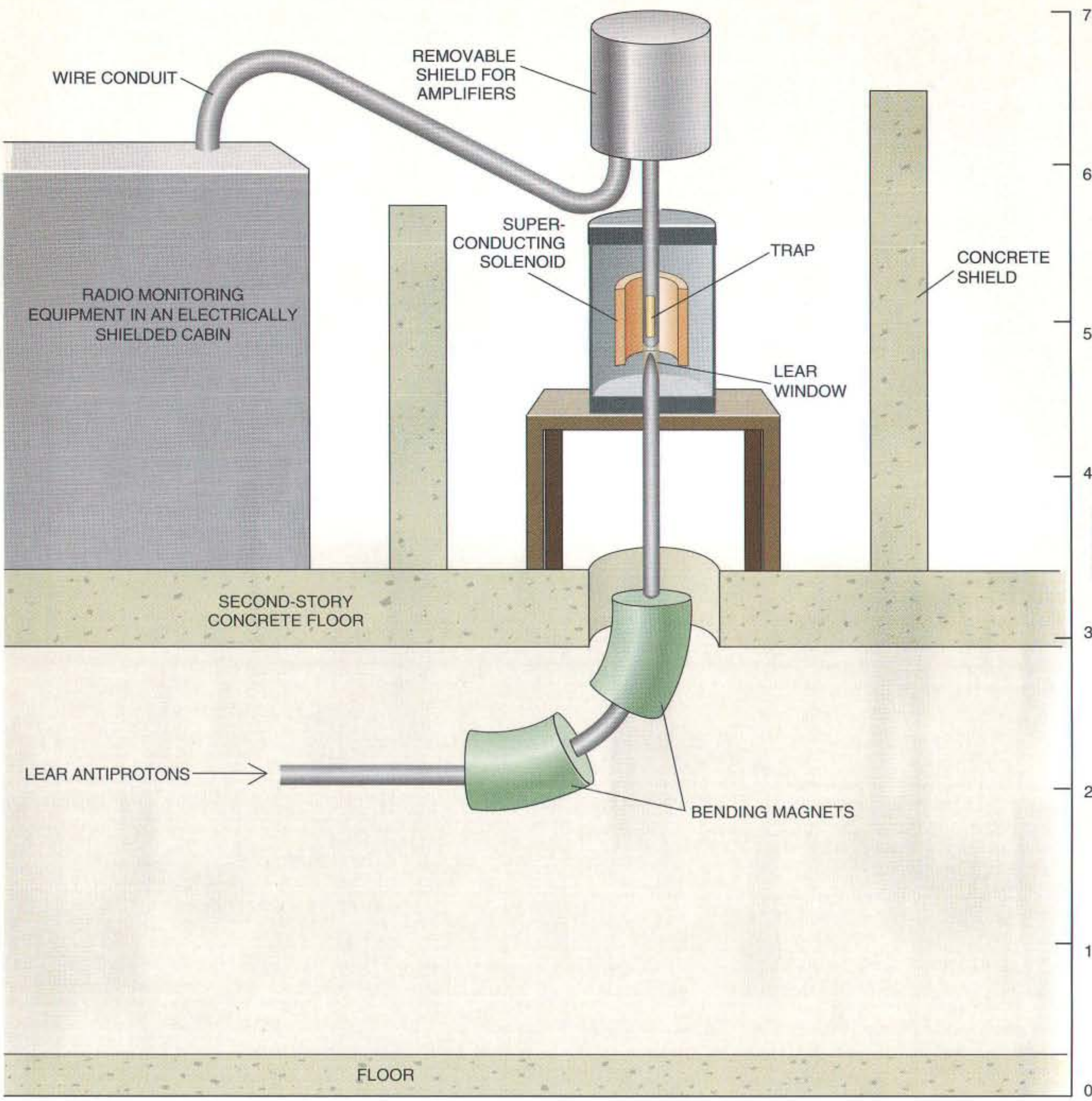
CERN enthusiasm now replaced CERN ambivalence. A semipermanent connection to LEAR was constructed for our experiments when the machine was shut down for one year. Our present apparatus, which we installed in 1988, now sits on a platform 4.3 meters above the ground. At the heart of the antiproton cooler lies the ion trap, a stack of gold-plated copper rings located in a magnetic field [*see illustration on page 41*]. Within the trap, charged particles make circular orbits perpendicular to the direction of a six-tesla magnetic field. The field, which is approximately 10 times more powerful than that generated by a strong permanent magnet, is produced by sending 37 amperes of current through a 25-mile coil of superconducting wire. Once the current has been introduced in the coil, it flows with no resistance, in seemingly perpetual motion. No external power is needed. Additional correction coils make the magnetic field constant over the small volume that is to be occupied by the antiprotons.

Voltages applied to electrodes in the trap keep charged particles from escaping out the upper and lower ends of the device. Electrons are trapped in a small, cloudlike formation before any antiprotons enter the trap. Negative tens of volts applied to the ring electrodes on either side of the small trapping region repel the electrons toward the center of the trap, where they are confined. The electrons rapidly radiate their energy, typically in a tenth of a second, and cool to the four-kelvin temperature of the surrounding electrodes. The trap is now prepared for the antiprotons.

The antiprotons crash through the bottom electrode of the trap, arriving in an intense 6-MeV pulse from LEAR. Immediately they begin to lose energy in random collisions with the particles that make up the electrode. Some antiprotons slow to a stop within the electrode and eventually annihilate. Others emerge along the axis of the trap with an energy that exceeds 3,000 electron volts (3 keV). These strike the upper electrode and annihilate. The remaining antiprotons, whose energies are below 3 keV, are those we can trap. Their number is maximized, typically to one in 5,000 of the incident antiprotons, by carefully choosing the thickness of the electrode. The filtered antiprotons travel upward until, repelled by the negative



ENERGY THERMOMETER contrasts the extremely large range of energies at which antiprotons (and protons) are stored for study. Each demarcation is a factor of 10 lower in energy than the one above. The Large Hadron Collider (LHC), proposed for CERN, the European laboratory for particle physics, and the Superconducting Super Collider (SSC), proposed for Texas (both to use protons), are at the top, and the Low Energy Antiproton Ring (LEAR) at CERN is in the middle. The new low-energy frontier described in the article is at the bottom, one ten-billionth of the LEAR energies. The energy units are electron volts [*see box on opposite page*], and the corresponding temperatures are expressed in kelvins (K).



ANTIPROTON JOURNEY into the trap apparatus begins at the lower left of the illustration, and the particles are turned upward by two bending magnets. The antiprotons leave the LEAR vacuum and enter the vacuum of the antiproton trap through a pair of windows located within the cold apparatus, which sits on a "tabletop" 4.3 meters above the ground.

voltage of the upper electrode, they turn around and head back. To prevent their escape, the "entrance window" (bottom electrode) is "slammed shut" by applying a –3,000-volt potential to it in less than 20 billionths of a second.

The captured antiprotons oscillate back and forth along the 12-centimeter length of the trap, passing through the cold, trapped electrons. Just as a heavy bowling ball would ultimately be slowed by collisions with light Ping-Pong balls, virtually all the antiprotons cool to thermal equilibrium with the trapped electrons in less than two minutes. (Electrons are the ideal cooling agent insofar as they cannot annihilate the precious antiprotons.) Typically 10,000 antiprotons from a single LEAR pulse are cooled in the small trap. Once the antiprotons are cooled, the electrons are allowed to leak out by selectively heating them with radio waves, as we temporarily reduce the confining voltages on the trap electrodes. We have observed no loss of extremely cold antiprotons, even when the particles were held for two months.

An immediate consequence of the long-term storage is that we have shown the antiproton to be stable for at least 3.4 months. Although notably less than the 10^{25}-year proton lifetime limit, our figure stands as the longest direct determination of the antiproton's lifetime. It could be made because there are fewer than 100 background gas atoms per cubic centimeter in the trap. This is a remarkably low pressure limit

(5×10^{-17} torr), a millionth of what can be measured by commercial vacuum gauges. Our high vacuum is achieved because background gas atoms stick to the cold electrodes within a tightly sealed container.

Once we had successfully stored the extremely cold antiprotons, we were able to measure and compare the mass of that particle with the mass of the proton 1,000 times more accurately than had previously been possible. Our effort was aided by Xiang Fei, Luis A. Orozco, Steve L. Rolston and Robert L. Tjoelker, from my research group at Harvard University, and by Johannes Haas of the University of Mainz.

The measurements are based on the fact that the "cyclotron" frequency of the circular orbit of a charged particle in a magnetic field is simply the product of the charge of the particle and the strength of the magnetic field, divided by the mass of the particle. In other words, a massive particle orbits more slowly than a light particle does. In the strong magnetic field we use, antiprotons and protons make approximately 90 million revolutions per second. We detect the radio signal emitted by the rapidly orbiting particles and measure the cyclotron frequency, which is 90 million cycles per second (90 MHz), by means of an FM radio receiver. We found that the antiproton and proton orbital frequencies were the same to within four parts in 100 million.

Much of the experimental effort goes into evaluating and reducing the uncertainties, taking into account the effects of the somewhat more complicated orbital motion actually exhibited by the particles in the trap. Because we made sure that the magnetic field did not change between the measurements, the charge-to-mass ratio of the antiproton and proton is shown to be the same to within four parts in 100 million. If both particles are assumed to have the same amount of charge, the mass of the antiproton is the same as the mass of the proton to within the same limit.

Holding the magnetic field constant during the measurements is especially difficult because the magnets in the nearby particle accelerator are turned on and off every 2.4 seconds. Fortunately, my graduate student Joseph N. Tan and I discovered we could design a superconducting solenoid, or wire coil, that senses changes in the external magnetic field and adjusts its own magnetic field to cancel those changes. The solenoid, which also supplies the strong magnetic field needed for our measurements, reduces fluctuations by a factor of 156. Our invention, now patented because of likely applications for magnetic resonance imaging and ion cyclotron resonance mass spectroscopy, illustrates the interplay between pure science and technology. The pursuit of fundamental physics goals pushes technology so hard that practical applications emerge.

In the near future we hope to measure the orbital frequencies of the antiproton and proton even more accurately. Several collaborators have now replaced graduate students and postdoctoral fellows who have moved on: Wonho Jhe, David Phillips and Wolfgang Quint, from my group at Harvard, and Julian Gröbner of the University of Mainz. Our early work has been promising—we have already increased the precision of our measurement by an additional factor of 40. We are also looking into the possibility of measuring the magnetic moment of the antiproton. The particle acts like an extremely small bar magnet; the magnetic moment is the effective strength of this magnet.

More accurate comparisons of antiprotons and protons will be difficult in the environment of a particle accelerator, and so it may become necessary to move antiprotons in our tabletop apparatus to a nearby location. In a similar apparatus, Harvard graduate student Ching-Hua Tseng and I recently transported trapped particles more than 3,000 miles across the U.S., from California to Nebraska and then from Nebraska to Massachusetts.

Comparisons of the orbital cyclotron frequencies of antiprotons and protons test the *PCT* invariance theorem. Historically, *P*, which stands for parity, was examined first. To understand the concept, imagine conducting an experiment in which the outcome is watched in a mirror. Now suppose a second experiment is constructed that is the mirror image of the first. If parity is conserved, the outcome of the second experiment should be identical to the outcome observed as the mirror image of the first experiment performed.

Until 1956, it was believed that reality was invariant under such a parity transformation. Early that year, however, Tsung-Dao Lee and Chen Ning Yang, then at Columbia University and the Institute for Advanced Study in Princeton, N.J., respectively, realized that the invariance of parity in weak interactions,

FIRST ANTIPROTON TRAP consisted of simple copper electrodes that were separated by glass spacers.

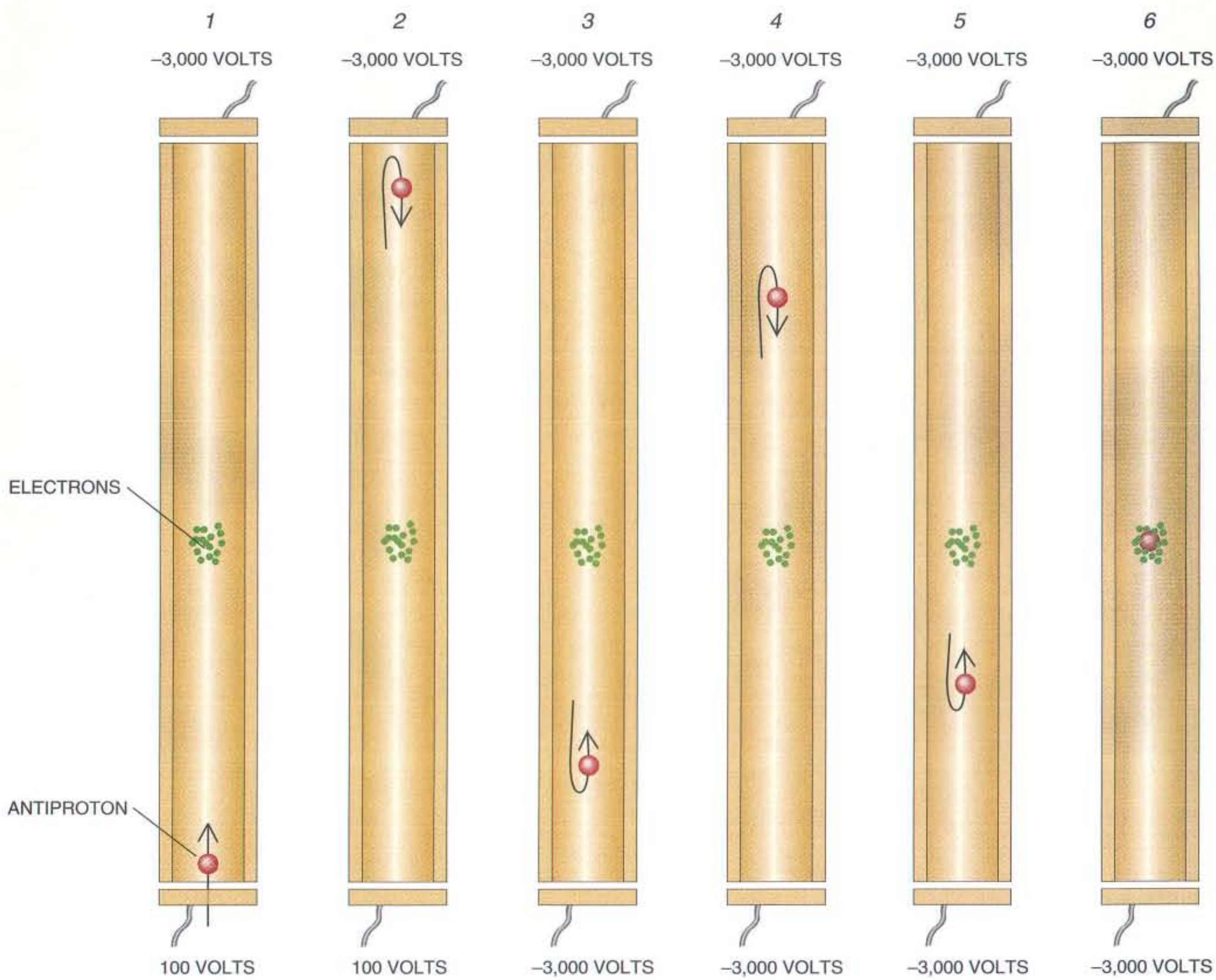
which are responsible for radioactive decay, had not yet been tested. Later that year Chien Shiung Wu and her colleagues at Columbia showed that mirror-image experiments did not produce mirror-image results when weak interactions were involved. The widespread belief in parity conservation was shattered.

Faith in a new invariance, *PC*, rapidly replaced the discredited notion. *C* stands for charge conjugation, a "thought experiment" process that turns particles into their corresponding antiparticles. To test whether *PC* is conserved, a mirror-image experiment is constructed, and all the particles in the experiment are replaced with their corresponding antiparticles. In 1964 James Cronin and Val L. Fitch, then at Princeton University, used particles called kaons to demonstrate, explicitly and unexpectedly, that *PC* is not conserved [see "A Flaw in a Universal Mirror," by Robert K. Adair; SCIENTIFIC AMERICAN, February 1988].

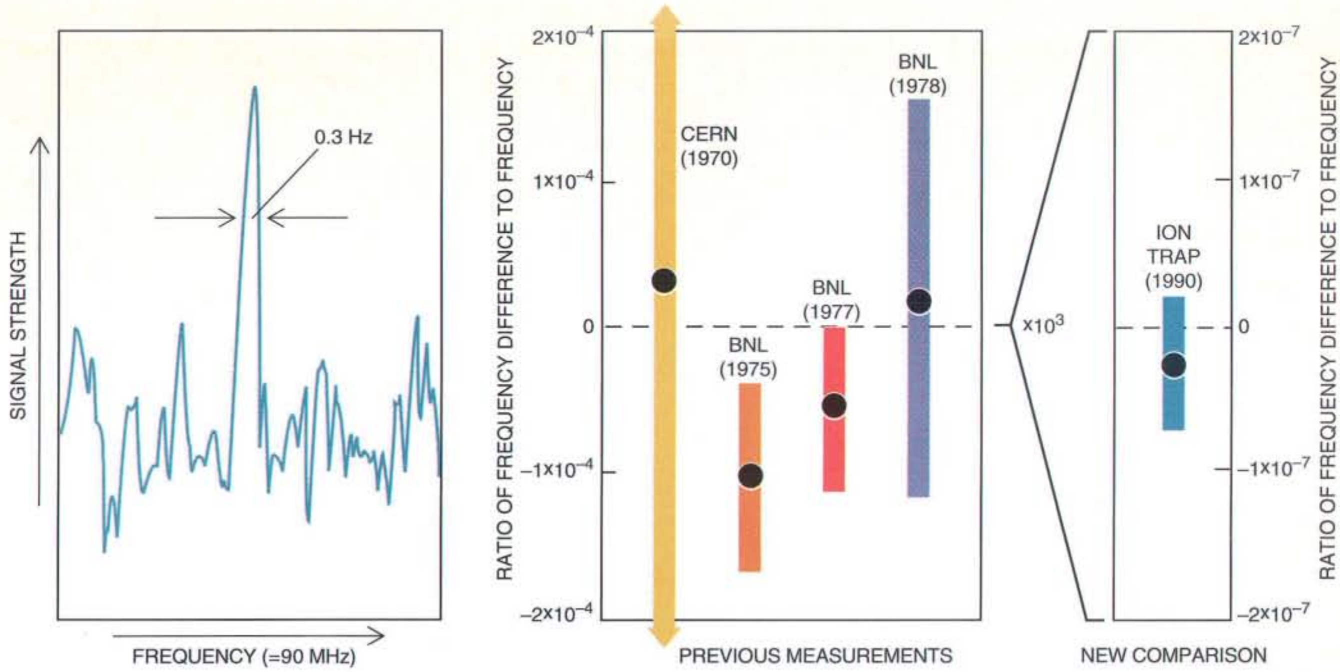
Today most physicists believe that *PCT* is invariant (the *T* stands for time reversal). Thus far theorists have yet to construct a reasonable theory in which *PCT* is not conserved. To test the invariance of *PCT*, imagine making a movie of an experiment's mirror image in which all the particles have been replaced by their corresponding antiparticles. Then a second experiment is performed to mimic what one sees in the film when it is run backward—when "time is reversed."

A consequence of *PCT* invariance is that the circular cyclotron frequencies of the antiproton and proton in a magnetic field should be identical. Our comparison thus tests *PCT* invariance and establishes that violations are smaller than the experimental uncertainties. Our experiment is currently one of the most accurate tests of *PCT* invariance. As our accuracy increases, we shall see whether this invariance of *PCT* continues to hold.

In the more distant future, extremely cold antiprotons should make it possible to produce and study antihydrogen, the antimatter atom formed by a positron in orbit around an antiproton. (Positrons, or antielectrons, are produced naturally in the radioactive decay of atomic nuclei.) Under the proper conditions, small numbers of



ANTIPROTON CAPTURE begins as the particle (*red*) leaves the entrance-window electrode within which it was slowed (*1*) and travels upward until it is repelled by -3,000 volts on the upper electrode (*2*). Reversing its direction, the antiproton travels downward toward the entrance window, until it is repelled there because the voltage on this window has in the meantime been changed from +100 to -3,000 volts (*3*). (The entrance window is initially held at +100 volts to prevent large numbers of electrons liberated from the electrode during the passage of antiprotons through it from entering the trap.) The antiproton is thus trapped, oscillating up and down between the two repulsive voltages. Repeated collisions (*4, 5*) with the electrons (*green*) in the trap cool the antiproton until it resides within the electron cloud (*6*).



ANTIPROTON RADIO SIGNAL (*left*) from the circular cyclotron motion of trapped antiprotons is strongest at the cyclotron orbit frequency. The ability to measure the difference between the circular frequencies of protons and antiprotons in a strong magnetic field is improved by a factor of 1,000 by using extremely cold antiprotons (*right*). Earlier measurements (*center*) made at CERN and at Brookhaven National Laboratory (BNL) were unable to reach this level of accuracy.

cold antihydrogen atoms should be formed by mixing large numbers of extremely cold antiprotons with large numbers of extremely cold positrons. In 1986 I thus outlined a program to make cold antihydrogen atoms and to confine them by their magnetic moments in a trap for neutral particles. It may also be possible to make and capture antihydrogen ions, each of which would consist of two positrons bound to an antiproton.

Several important experiments could be performed on trapped antimatter atoms. Comparisons of the internal oscillation frequencies of antihydrogen and hydrogen would test *PCT* invariance even more accurately. It might also be possible to measure directly the gravitational properties of the antimatter atom, which would be electrically neutral and hence not extraordinarily sensitive to stray electrical forces.

Antihydrogen production is an ambitious and difficult undertaking that will take some time to realize. Estimated production rates are very low. Techniques must be devised to cool antihydrogen to the low energies required for trapping—conventional cooling methods involve collisions with cold surfaces, which would cause antihydrogen to annihilate. It also remains to be shown that accurate spectroscopic measurements can be done with only a few atoms in a trap. One very encouraging circumstance is that antihydrogen is more easily detected than hydrogen. Pions emitted on annihilation record the presence of a single antihydrogen atom.

Contemplated antihydrogen production requires the largest possible number of cold antiprotons. To this end, we have demonstrated that we can "stack" antiprotons harvested from successive LEAR pulses. Instead of ejecting the cold electrons from the trap once the antiprotons from one pulse have been stored, the electrons are used to cool the antiprotons from many successive pulses. In this way, during one hour, more than 100,000 cold antiprotons have been stacked, or added to one another, in the trap. We estimate that our current apparatus is capable of capturing and cooling up to one million antiprotons. It should be possible to employ a larger trap and higher trapping voltages to capture and cool an even larger number of antiprotons.

It is always difficult to predict what will transpire at a frontier just beginning to be explored. Whatever experiments are done with cold antiprotons, however, they are likely to be small, tabletop investigations especially suited for students to carry out as part of their training. Perhaps the antiprotons will even be transported away from their source. At present, we skim only a small fraction of the available antiprotons from the huge high-energy experiments for which the antiproton sources were constructed. In the future, large, high-energy experiments will use protons instead of antiprotons. A major challenge will be to retain access to an antiproton source so that work with cold antiprotons may continue.

FURTHER READING

FIRST CAPTURE OF ANTIPROTONS IN A PENNING TRAP: A KILOELECTRONVOLT SOURCE. G. Gabrielse, X. Fei, K. Helmerson, S. L. Rolston, R. Tjoelker, T. A. Trainor, H. Kalinowsky, J. Haas and W. Kells in *Physical Review Letters,* Vol. 57, No. 20, pages 2504–2507; November 17, 1986.

GEONIUM THEORY: PHYSICS OF A SINGLE ELECTRON OR ION IN A PENNING TRAP. L. S. Brown and G. Gabrielse in *Reviews of Modern Physics,* Vol. 58, No. 1, pages 233–311; January 1986.

ANTIHYDROGEN PRODUCTION USING TRAPPED PLASMAS. G. Gabrielse, S. L. Rolston, L. Haarsma and W. Kells in *Physics Letters,* Vol. 129, No. 1, pages 38–42; May 2, 1988.

COOLING AND SLOWING OF TRAPPED ANTIPROTONS BELOW 100 meV. G. Gabrielse, X. Fei, L. A. Orozco, R. L. Tjoelker, J. Haas, H. Kalinowsky, T. A. Trainor and W. Kells in *Physical Review Letters,* Vol. 63, No. 13, pages 1360–1363; September 25, 1989.

THOUSANDFOLD IMPROVEMENT IN THE MEASURED ANTIPROTON MASS. G. Gabrielse, X. Fei, L. A. Orozco, R. L. Tjoelker, J. Haas, H. Kalinowsky, T. A. Trainor and W. Kells in *Physical Review Letters,* Vol. 65, No. 11, pages 1317–1320; September 10, 1990.

Directed Molecular Evolution

Biochemists have harnessed Darwinian evolution on a molecular scale. Through cycles of selection, amplification and mutation, populations of macromolecules can be pushed to evolve toward any functional goal

by Gerald F. Joyce

Inspired by the accomplishments of Darwinian evolution in nature, scientists are now beginning to take evolution into their own hands. Evolution is being carried out in the laboratory, not at the level of organisms or even at the level of cells but at the level of individual macromolecules. The products of these experiments in evolution are molecules exhibiting properties that conform to the demands of the experimenter.

In some ways, directed evolution is like the captive breeding programs practiced by horticulturists, cat fanciers and the like. If one wants a redder rose or a fluffier Persian, one chooses as breeding stock those individuals that best exemplify the desired characteristic. So, too, if one wants a molecule that exhibits a particular chemical trait, then one selects from a large population of molecules those individuals that best manifest the property. From these selected individuals, the experimenter generates progeny molecules that to varying degrees resemble their parents. This process of selection and not quite faithful duplication is repeated until the desired end is achieved.

The power of directed evolution lies in the strength of large numbers. It is not unusual for a biochemist to survey 10^{13} different molecules at a time (imagine doing that with cats!). A generation of molecules may last only one or two days, from the time they are selected until the time they are used to create more molecules. Selection can be very stringent: if merely one molecule in a billion exhibits a serviceable property, then it is sufficient to act as "breeding stock" for the next generation.

GERALD F. JOYCE is an RNA biochemist with a long-standing interest in molecular evolution. After studying at the University of Chicago, he went on to earn his medical degree and doctorate from the University of California, San Diego, in 1984. He held a postgraduate medical position at Mercy Hospital and Medical Center and then a postdoctoral research position at the Salk Institute for Biological Studies, both in the San Diego area. In 1989 he joined the faculty of the Scripps Research Institute in La Jolla, Calif., where he is affiliated with the department of chemistry and the department of molecular biology.

A molecular biologist can look at an organism's genes and read them as a kind of historical document about its evolution—a document written not in ink on paper but as chemical components in a DNA molecule. A biochemist can similarly look at the evolving molecules in an experiment and obtain precise information about their characteristics at any point in their history. That information can in turn guide further explorations. Directed evolution promises to yield entirely new classes of reagents for industrial applications and novel drugs for medicine. It may even help humankind keep pace with the relentless evolution of disease.

Darwinian evolution fundamentally involves the repeated operation of three processes: selection, amplification and mutation. Selection, whether it occurs naturally or artificially, is a winnowing process that separates the "haves" from the "have-nots." In nature the haves among organisms are those that survive to reproductive age, find a suitable mate and produce viable offspring. In the laboratory the haves are molecules that meet whatever criterion is imposed by the experimenter. For example, one might be interested in molecules that bind tightly to a certain toxin. Molecules that bind to it are retained, and those that do not are discarded.

Amplification is the process of generating progeny or, more precisely, making copies of the inherited genes. In nature, selection and amplification are coupled directly: when organisms choose their mates, the act of procreation results in offspring that carry copies of their parents' genes. In the laboratory the experimenter must link selection to amplification by allowing only the molecules that meet imposed criteria to have progeny. To be precise, it is not necessarily the selected molecules themselves but a genetic description of them that must be amplified.

Mutation introduces variation, without which there is no hope of evolutionary progress because there are no alternatives on which selection can act. Many laboratory systems introduce variation at the outset of a selection experiment to create a heterogeneous population of molecules. The most desirable ones are then reaped from the population through rounds of selection and amplification.

Strictly speaking, such a process is not evolutionary, because it does not include an opportunity for new mutations to arise in each generation. Some of the more sophisticated laboratory systems, however, provide variation in every generation, so that selection, amplification and mutation can operate steadily and in concert.

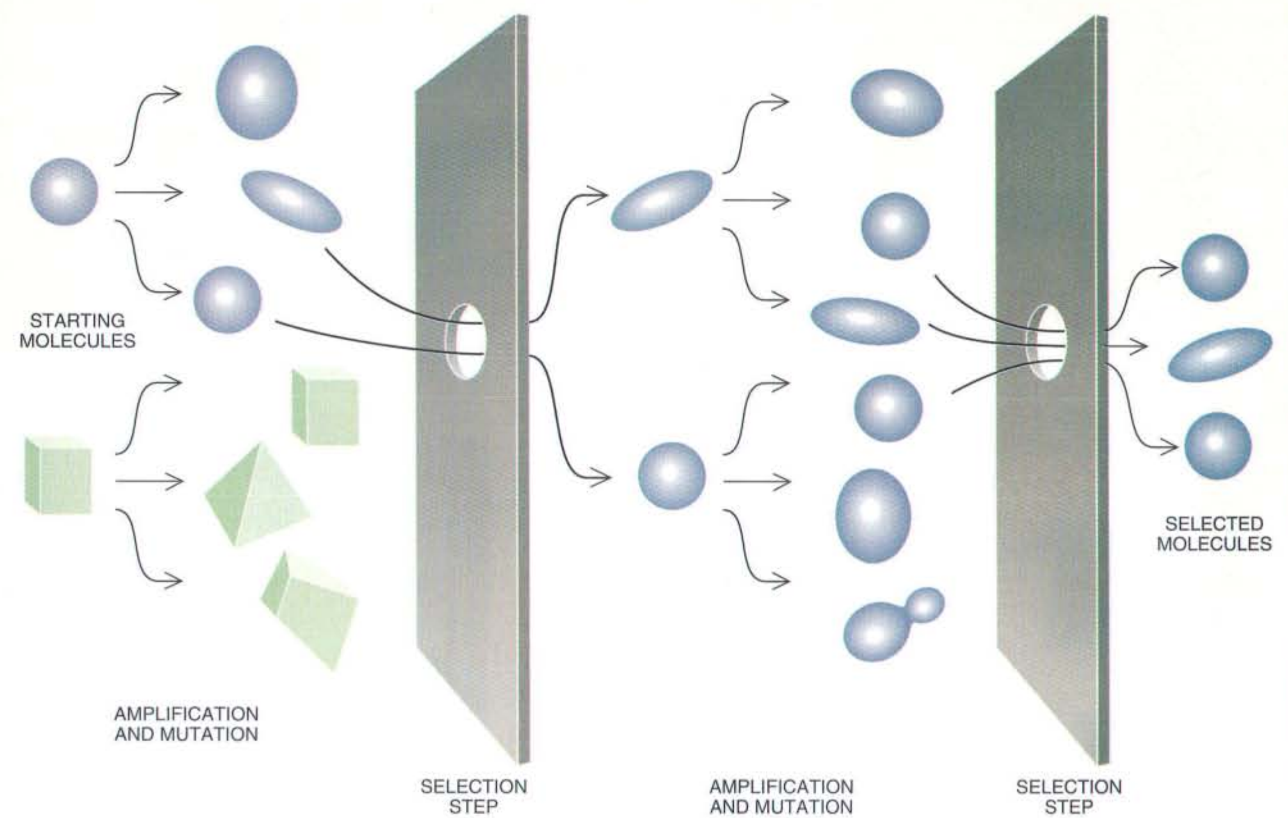
Darwinian evolution in the laboratory was first demonstrated in the late 1960s by Sol Spiegelman of the University of Illinois and his co-workers. Spiegelman had been studying a protein of the Qβ bacteriophage, a small virus that infects the intestinal bacterium *Escherichia coli*. The strand of RNA that serves as genetic material for the bacteriophage contains only four genes, one of which encodes a protein enzyme called a replicase. The replicase is essential for the survival and proliferation of the virus because it makes copies of the viral RNA genome.

Spiegelman knew that if he mixed the viral RNA with its replicase protein and added ribonucleoside triphosphates (the building blocks of RNA), the replicase would make additional copies of the viral RNA. In this way, he could carry out amplification of a genetic molecule in a test tube. The amplification process has a built-in mutation feature

DIVERSITY is crucial for evolution. Experimenters can screen large heterogeneous populations by selecting for individuals that precisely fit any functional requirement they choose. Through the selection, amplification and mutation of diverse macromolecules, biochemists are now beginning to direct the evolution of new drugs and reagents.

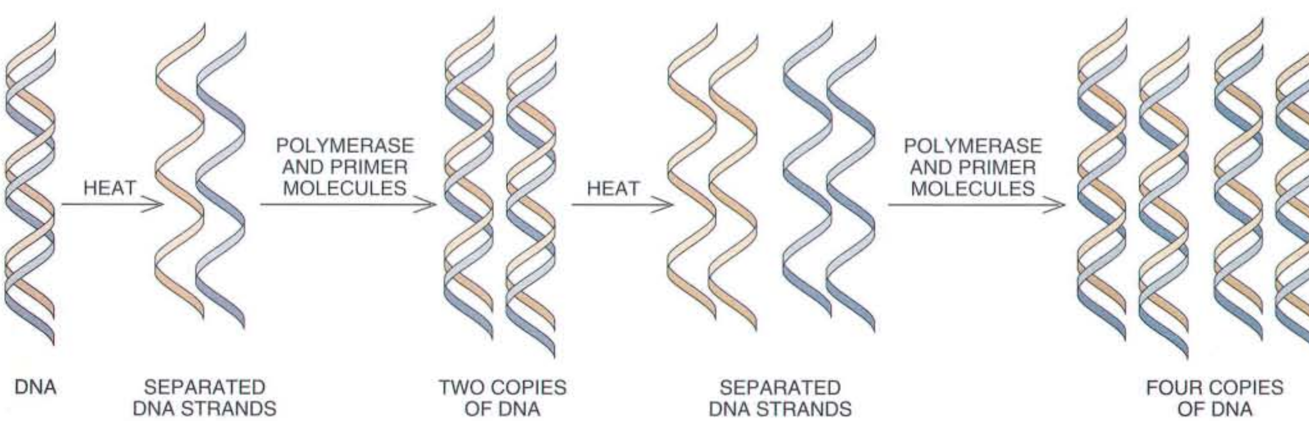
How Directed Evolution Works

A "molecular obstacle course" is the metaphor for directed evolution: macromolecules compete to clear functional hurdles imposed by the experimenter. Those molecules selected in this way are then amplified, or copied, to produce a new generation that generally resembles its parents but also contains mutations. Successive rounds of selection and amplification with mutation create a population of molecules with a desired trait.



because the viral replicase protein is a sloppy worker: it almost always makes one or two errors when it copies a sequence of RNA. As a selection criterion, Spiegelman chose the following question: What will happen to the RNA molecules if the only demand made on them is the Biblical injunction *multiply,* with the biological proviso that they do so as rapidly as possible?

Putting together the selection, amplification and mutation features of the Qβ system, Spiegelman conducted what is known as a serial transfer experiment. He allowed the Qβ replicase to amplify Qβ RNA in a test tube for 20 minutes. During this time, the replicase made copies of the parent molecules as well as copies of the progeny, committing occasional errors in the process. Spiegelman then transferred a sample of the



POLYMERASE CHAIN REACTION is a technique used in directed evolution to make copies of specific DNA molecules. If a double helix of DNA is heated to separate its strands, enzymes called polymerases (in conjunction with DNA subunits and short primer sequences) can make new copies of the original helix. This process can be repeated endlessly.

reaction mixture to another test tube containing a fresh supply of the replicase protein and the four nucleoside triphosphates. The cycle of amplification and sample transfer was repeated 74 times. In each cycle the procedure favored the proliferation of those molecules that gave rise to the most progeny before the transfer step.

Periodically, Spiegelman would tighten the screws of selection by reducing the time available for amplification. Molecules that required a longer time to make copies of themselves would then not be able to leave as many progeny as molecules that required a shorter time. Indeed, as evolution proceeded in the test tubes, the surviving RNAs became progressively shorter because smaller molecules could be copied more rapidly and thus more frequently than the full viral genome. By the 74th transfer the evolving RNA molecules had jettisoned 83 percent of the original Qβ genome, retaining only the part that the replicase needed to do its job. Although these molecules were no longer competent to infect *E. coli,* they had succeeded quite nicely in doing what Spiegelman had asked of them: they had increased their replication rate roughly 15-fold.

The Qβ system has contributed significantly to our understanding of Darwinian evolution at the molecular level. Since Spiegelman's work, several investigators have used the Qβ evolution system to produce interesting RNA molecules. For example, through directed evolution, Leslie E. Orgel of the Salk Institute for Biological Studies in San Diego and his co-workers created variants of Qβ RNA that are resistant to the drug ethidium bromide, which usually inhibits the replication of RNA in the test tube.

The Qβ system has one serious limitation, however: the replicase protein is very finicky about which RNA sequences it will replicate. Consequently, the overriding selection constraint on the Qβ system—one far more pressing than any constraints imposed by the experimenters—is that the evolving RNAs remain good substrates for the replicase protein. The situation is not unlike that of a parent telling a child: "You can do anything you want with your life, so long as you stay home to run the family business."

The choosiness of the Qβ replicase restricts the versatility and hence the usefulness of the system. In recent years, however, researchers have developed systems for evolution in the laboratory that keep the replication and selection processes separate and are therefore more flexible. This division became possible with the advent of amplification procedures that, unlike the Qβ replicase, are indifferent to the sequence of the gene.

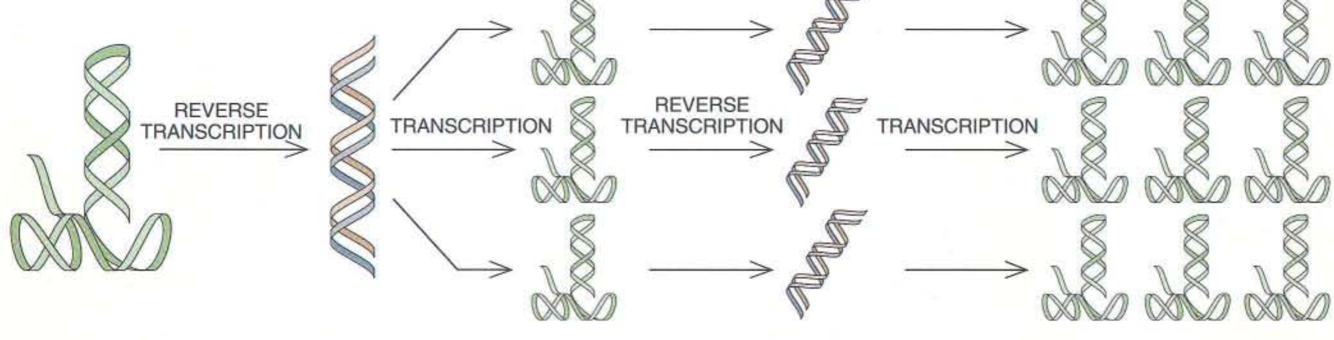
One such procedure is the polymerase chain reaction, which allows a millionfold amplification of a DNA sequence in a few hours [see "The Unusual Origin of the Polymerase Chain Reaction," by Kary B. Mullis; SCIENTIFIC AMERICAN, April 1990]. Polymerases are enzymes that assemble nucleotides into new DNA or RNA strands complementary in sequence to an original template strand. (The Qβ replicase is one type of polymerase.) The polymerase chain reaction relies on the fact that DNA strands synthesized by polymerases can themselves serve as templates for further rounds of replication.

Thus, polymerases copy each strand of a DNA double helix containing a gene to produce a second double-helix molecule. They then copy both the old and the new double helices to produce a third and a fourth; all four helices are then copied to make eight, and so on. The number of helices grows geometrically, so that within just a few hours the original DNA gene has been duplicated millions of times.

A related general-purpose amplification technique can rapidly make copies of RNA strands. It involves two polymerase proteins, one to copy RNA to DNA (that is, to make a DNA strand complementary to an RNA template) and the other to copy the DNA back to RNA. The result is the same as in the polymerase chain reaction: by repeatedly making copies of the copies, one quickly ends up with an astonishing number of identical RNA genes. Using the RNA amplification procedure, one can obtain millions of copies in an hour. If greater replicative horsepower is needed, RNA amplification can be made to operate in tandem with the polymerase chain reaction to produce billions of copies of a gene.

Ironically, as originally developed, the polymerase chain reaction and the RNA amplification system shared a drawback: they were too accurate. That is, unlike the Qβ system—in which copying mistakes occurred on the fly, as it were, during amplification—the polymerase chain reaction and the RNA amplification system were not sufficiently susceptible to error to meet the mutation requirements of a directed evolution endeavor. Fortunately, modified versions of these procedures now allow mutations to occur at a reasonable and controllable frequency.

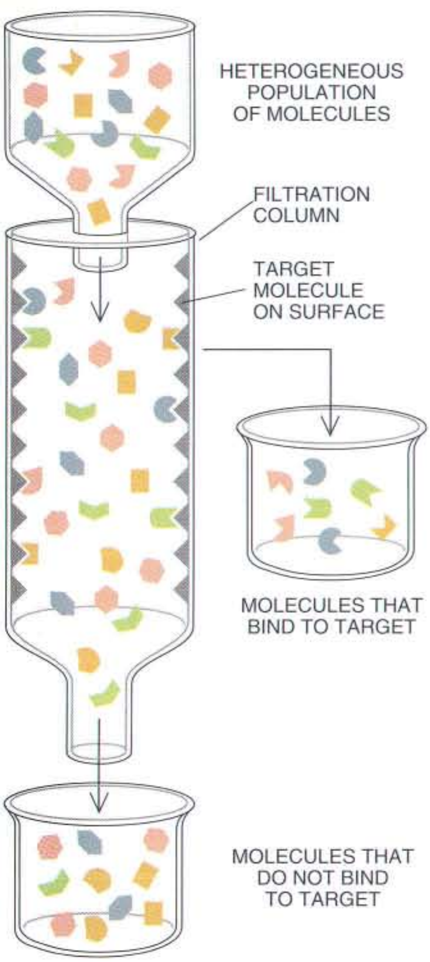
With these powerful gene amplification procedures in hand, biochemists are now free to explore a wide range of selection constraints. The genetic molecules DNA and RNA are particularly useful targets for further explorations because the base sequences that determine their chemical and physical properties are synonymous with their ge-



RNA AMPLIFICATION is the copying technique used in directed evolution if the molecules under study are made of RNA. First, a reverse transcription enzyme makes a DNA copy of an RNA molecule. A transcription enzyme then makes multiple copies of the original RNA from the DNA. The process repeats itself automatically at a constant temperature.

netic makeup. They can be selected on the basis of their function, then amplified by copying the genetic sequence responsible for that function. Normally, one thinks of proteins as the premier functional macromolecule—after all, nature employs proteins as the primary agents of catalytic function in cells. Nevertheless, as Thomas R. Cech of the University of Colorado and Sidney Altman of Yale University first established with their Nobel Prize-winning research, RNA can be a catalyst. It is reasonable to assume that DNA could be a catalyst as well. Other vital chemical activities, such as binding to target molecules, are also well within the province of DNA and RNA.

Biochemists, and perhaps those involved in biomedical research in particular, have strong incentives for trying to develop DNA and RNA as applied chemical agents. Today if a middle-aged man were wheeled into an emergency room after suffering a heart attack, he could be treated with the state-of-the-art drug streptokinase, a bacterial protein that dissolves blood clots. Unfortunately, he might suffer an allergic reaction to the drug—a not uncommon occurrence when someone is injected with a foreign protein. With the proper care, the patient may survive both the heart attack and the allergy, but if he suffers another heart attack in the future, his treatment options will be limited. Anticoagulant drugs based on designer DNA or RNA might someday be able to offer relief without provoking hypersensitivity to streptokinase.



SIFTING THROUGH MOLECULES by passing them through a filtration column is one method of selection. The molecules that have an affinity for the target stick to the surface and can be recovered later; the rest wash into a waste container.

Directed evolution is already helping the biotechnology industry make such drugs a reality. John J. Toole and his colleagues at Gilead Sciences wanted to develop an inhibitor to the clot-forming protein thrombin. Rather than look to nature for an answer, they decided to survey DNA molecules for useful properties. They constructed 10^{13} single-strand DNA molecules, each with a different sequence, and exposed this DNA population to thrombin that had been immobilized on a solid surface. (To guard against unwanted, incidental adhesion that was not specific to thrombin, Toole's group used molecules that in previous tests had shown no affinity for the surface alone.)

Nearly all the DNA molecules failed to bind to the thrombin and were easily washed off the surface. Yet 0.01 percent of them, about 10^9 molecules, remained bound. After recovering this fraction, Toole and his co-workers used the polymerase chain reaction to create 10^{13} progeny from this selection of molecules.

The team repeated this process of selective amplification for five generations. The population of DNA molecules they ultimately produced was highly enriched for the ability to bind thrombin. Molecules isolated from the population bound tightly to thrombin and, as a consequence, could inhibit the formation of clots. These molecules are now being tested in baboons and monkeys, in which they have been shown to act as effective anticoagulants.

Thrombin is not a nucleic acid-binding protein, yet a DNA sequence was found that could bind thrombin avidly. The existence of this sequence could not have been predicted a priori, but, given the chance, it declared itself through repeated rounds of selective amplification.

Large populations of DNA or RNA molecules have been surveyed for the ability to bind to a variety of targets. Many of the early successes involved targets that were already known to interact with nucleic acids, such as the regulatory proteins that bind specific RNA sequences inside cells. To learn about the interactions between the RNA and the protein targets, biochemists traditionally had to make single, deliberate mutations in the RNA and watch how they affected the interaction. In this way, one could gradually piece together the major determinants of the binding mechanism by playing a molecular game of 20 Questions.

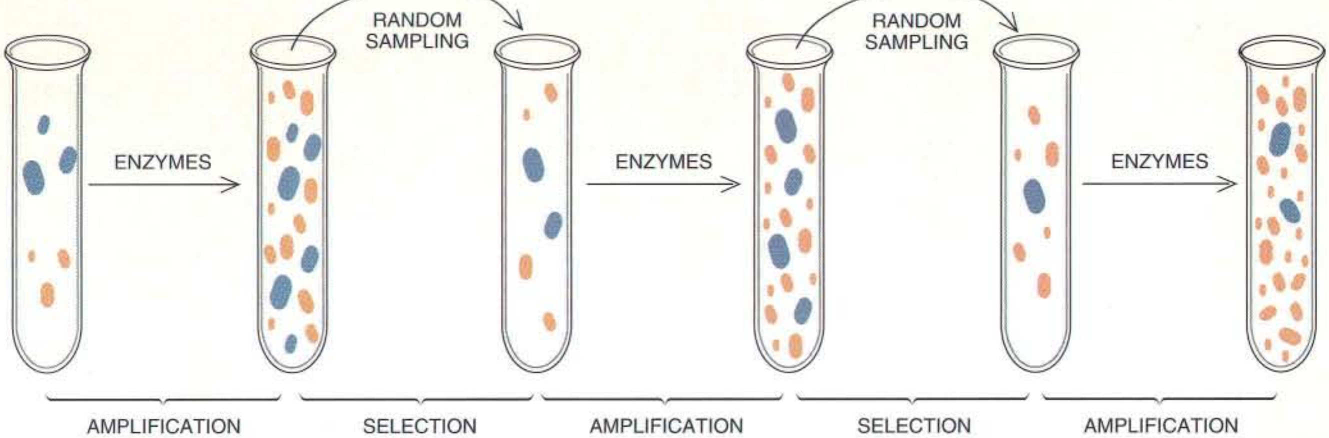
Today the tendency is to make 10^{13} mutations at one time and let selective amplification do the work. (Larger starting populations can be and sometimes are produced, but the cost of synthesizing more than 10^{15} molecules becomes prohibitive.) First, all or part of the RNA sequence is randomized—that is, randomly chosen nucleic acids are substituted for the original subunits in the RNA molecule. The experimenter sifts through the resulting heterogeneous population of RNAs to find the sequences that best bind to the protein target. The sifting is done in much the same way that Toole did it in his search for thrombin-binding molecules: by exposing the population to proteins attached to a surface.

Finally, in the amplification step, the selected molecules are reverse-transcribed into DNA sequences, which can be amplified many times over by the polymerase chain reaction. The DNA molecules can then be transcribed into RNA. Alternatively, the selected RNAs could be copied directly by the RNA amplification procedure.

The target for DNA or RNA binding could be any type of molecule, not just a protein. One of the first successful selective amplification experiments was conducted in 1990 by Andrew D. Ellington and Jack W. Szostak of Harvard Medical School, who used small organic dyes as the target. They screened 10^{13} random-sequence RNAs and found molecules that bound tightly and specifically to each of the dyes.

Recently they repeated this experiment using random-sequence DNAs and arrived at an entirely different set of dye-binding molecules. When the selected DNA sequences were transcribed into RNA molecules, they did not bind to their dye targets, which suggests that the DNA and RNA molecules bound to the dyes by quite different mechanisms.

That observation reveals an important truth about directed evolution (and indeed, about evolution in general): the forms selected are not necessarily the best answers to a problem in some ideal sense, only the best answers to arise



PROLIFIC MOLECULES can be selected in serial transfer experiments. Molecules are exposed to enzymes and given time to replicate; a small sample is transferred to a fresh test tube, and the process is repeated. Because the time that a molecule needs to replicate depends on its size, small ones (*orange*) can make more copies of themselves than can large ones (*blue*) in any period. Shorter molecules gradually come to predominate in the system as large ones disappear from the competition.

in the evolutionary history of a particular macromolecule.

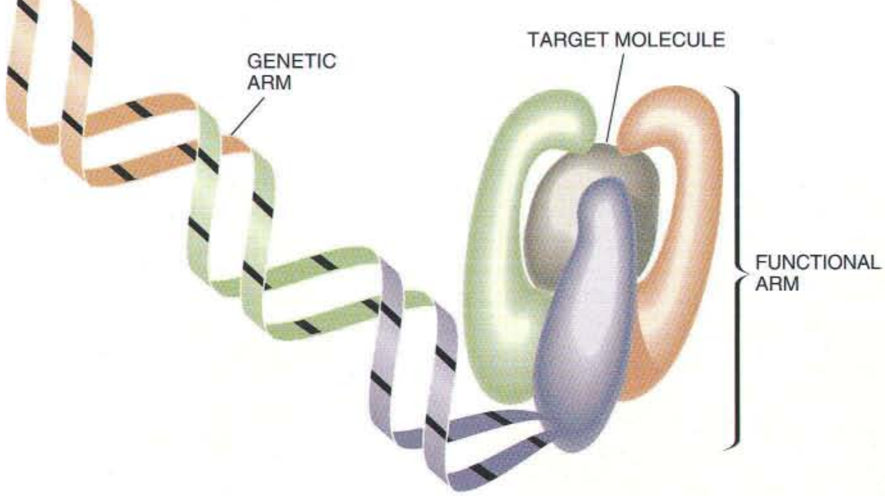
Directed evolution experiments need not be limited to DNA and RNA. In principle, any population of macromolecules is suitable for selection if researchers have an easy way to amplify the genetic descriptions of the individuals selected from it. Taking this broad-minded view, Sydney Brenner and Richard A. Lerner of the Scripps Research Institute in La Jolla, Calif., have invented their own language for genetic description that can be applied to virtually any macromolecule.

Brenner and Lerner's approach is to prepare dual-purpose molecules that have two "arms" joined at a common point. One arm is a functional macromolecule that can be tested for its ability to bind to a target. Its subunits might be amino acids, sugars or any other organic compounds. The other arm is a genetic macromolecule, typically DNA, that encodes a description of the functional arm. The sequence of nucleotides in the genetic arm acts as a record or bookkeeping device for listing the subunits in the functional arm.

The two arms are synthesized in parallel, first by adding a subunit to the functional arm, then by adding the appropriate nucleotides to the genetic arm. When the two arms are completed, the molecule is tested for its ability to bind to a target. The experimenter can recover the molecules that bind to the target and amplify their genetic arms with the polymerase chain reaction. By examining the sequence of the amplified DNAs, one can decode the composition of the functional arm of the selected molecules.

This selective amplification system illustrates the symbolic nature of genetic information. In preparing the dual-purpose molecules, the experimenter begins by choosing or making up a genetic code of nucleotides for describing the subunits in the functional arm. Brenner and Lerner have used three nucleotides to represent each functional subunit, just as nature uses triplets of nucleotides to specify the amino acid subunits in proteins. If an experimenter is feeling more verbose, however, four or more genetic symbols might be used to encode each functional subunit. As the number of different potential subunits in the functional arm grows larger, more genetic symbols will be needed to provide a vocabulary large enough to describe each subunit uniquely.

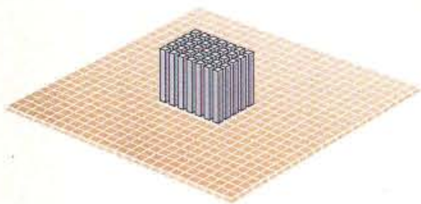
Regardless of what type of macromolecule one chooses to use in a directed evolution experiment, the initial step is the construction of a heterogeneous molecular population. There are three basic strategies for achieving this aim. The first is to prepare all the possible sequences of a given length. If a proposed macromolecule is to be 15 subunits long and there are four different types of subunits, the number of possible arrangements is four raised to the 15th power, or more than a billion. A starting population of 10^{13} molecules would therefore contain about 10,000 copies of each sequence. Some selective



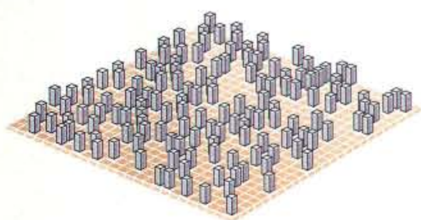
TWO-ARMED MOLECULES could become the highly diverse subjects of future directed evolution experiments. The functional arm, which consists of amino acids or other subunits, can be tested for its ability to perform a task, such as binding to a target. The genetic arm, which is made of DNA or RNA, contains a coded description of the functional arm's composition. This description can be replicated and analyzed.

amplification schemes, such as those involving Brenner and Lerner's dual-purpose molecules, require that the initial population be constructed according to such an all-possible-sequences strategy.

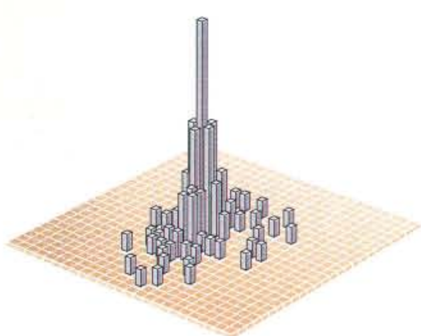
Yet as the length of a macromolecule grows, the number of possible variations on it increases exponentially. Consequently, in many experiments, it is not feasible to make or study all the variations. In those cases, a "shotgun" strategy for constructing a starting population may be preferable. The idea is to create a number of randomly different macromolecules that is extremely large but does not encompass all the possible arrangements of subunits.



BLANKET STRATEGY



SHOTGUN STRATEGY



FOCUSED STRATEGY

MUTATION STRATEGIES for designing the starting populations in directed evolution experiments can take several forms. In these figures the grid represents the range of macromolecular possibilities, and the height of the blocks is the relative abundance of one possibility in the population. In the blanket strategy, all the possible variations within defined limits are included. When the limits are less well defined, the so-called shotgun strategy ensures that a random sample of the possibilities is represented. The focused strategy uses a cluster of mutations resembling a molecule that has desirable properties.

For example, as Toole searched for thrombin-binding DNAs, he worked with molecules that had 60 randomized positions. A comprehensive population of all the possibilities would need to have had 4^{60} (roughly 10^{36}) different molecules—far too many for any researchers to have synthesized and screened. The population of 10^{13} molecules with which Toole actually started contained only a minute fraction of the possibilities, yet the successful outcome of the experiment demonstrated that it was still a large enough sample for directed evolution to shape it. When hunting for molecules with novel properties, it is often reasonable to adopt a shotgun approach.

On the other hand, a third, more focused strategy for building a starting population may be better if an investigator thinks a desired macromolecule is likely to be a variant of a known one. He or she can use the known molecule as a master sequence and introduce occasional mutations in a search for refinements or improvements. The investigator chooses a frequency of mutation that will ensure each molecule in the population contains a few differences from the master sequence; some molecules will of course contain more than others. The structured population that results from this approach contains large numbers of all the close relatives of the master sequence and an increasingly sparse sampling of sequences that are more distantly related.

Natural Darwinian evolution among organisms exploits the third strategy: nature introduces variations on the parental gene sequences, which are themselves the results of previous selection and amplification. Because the randomization of sequences occurs in every generation, heterogeneity is always maintained in the population. Iterative randomization, as this repetitive strategy is called, is especially powerful because it allows the selected mutants to give rise to still more mutants, some of which may prove to be even more advantageous than their predecessors. The one-time shotgun approach, in contrast, is useful for finding new leads but not for pursuing them to a desired end.

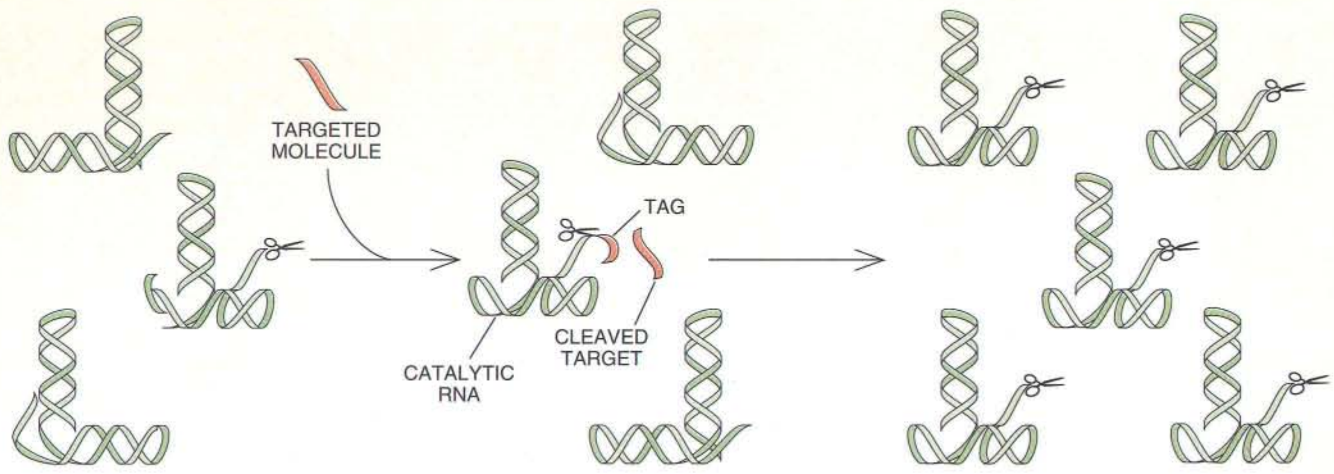
A subtle but important point about Darwinian evolution and the power of iterative randomization is worth noting. Because novel mutations augment the existing variation, the evolutionary search is biased—some observers would even say guided—by selection events that have already occurred. Evolution does not show foresight; rather the genes in the population at any given time reflect the properties that were advantageous in previous generations. Moreover, the number of copies of a particular genetic sequence is proportional to the selective advantage it confers. Within any extended family of related sequences, those sequences that are most advantageous and most abundant will spawn the largest number of progeny. New mutants are therefore generated preferentially within those branches of the family tree that have demonstrated an ability to generate advantageous mutants.

An iterative randomization strategy is the key to realizing the power of Darwinian evolution in the laboratory. The pieces are now in place not only to begin surveying large populations of DNA and RNA molecules for the ability to perform desired functions but also to push the evolution of the molecules to optimize their functional characteristics.

My own laboratory has learned to apply Darwinian evolution to virtually any population of RNA molecules. We are particularly interested in the so-called ribozymes, the RNA molecules that have catalytic function. Only a few ribozymes are known to exist in nature, and they account for only a limited repertoire of catalytic behaviors. We can pick up where nature has left off, however, through an accelerated process of evolution in the laboratory that directs ribozymes toward novel catalytic behaviors.

In one such study, we began with a naturally occurring ribozyme obtained from the protozoan *Tetrahymena thermophila*. This ribozyme has the ability to cut and splice targeted RNA molecules. We wanted to develop a ribozyme that would similarly cut a DNA target. Such a ribozyme could have therapeutic uses; for example, it might be used to cleave the unwanted DNA of an infecting virus inside a cell.

We generated a starting population of 10^{13} variants on the *Tetrahymena* ribozyme. We then exposed the population to a DNA substrate. Only a small fraction of the RNA molecules were able to cleave the substrate, but in so doing they incidentally became tagged with one piece of the cleavage product. The tagged individuals were selectively amplified, and new mutations were introduced during the amplification process to produce a second generation of molecules that had a greater ability to cleave DNA than did the starting population.



RNA THAT CUTS DNA was selected in experiments by the author. A heterogeneous population of RNA molecules with varied, unknown capabilities was exposed to a DNA target. The RNA molecule that had the ability to cleave the target did so; in the process it became tagged with a piece of the DNA. The tag enabled the author to amplify that RNA preferentially. Ten repetitions of this procedure created a population of RNAs with DNA-cleaving ability.

This entire cycle of events was repeated 10 times, resulting in a novel population of RNA molecules that were quite adept at cleaving DNA.

Darwinian evolution in the laboratory, unlike Darwinian evolution in nature, is a process we can literally hold in our hands. Passing generations need not become extinct; they can instead be stored in the freezer and revitalized at any time simply by thawing them out. Powerful techniques derived from recombinant DNA technology enable researchers to isolate individual molecules from any generation, to determine their entire genetic sequence and to measure precisely their catalytic properties. It is possible to reconstruct in detail the course of molecular evolution. One can even go back to any point in history and start over, using either the same selection criteria or new ones. Darwin's game of repeated selection, amplification and mutation has become the scientist's plaything.

The technology of directed evolution is still in its infancy. More work needs to be done to increase the size of the molecular populations that can be manipulated, to increase the stringency of the selection schemes and to decrease the amount of time required to create each generation. Darwinian evolution—that is, selective amplification coupled to an iterative randomization strategy—is at present applicable only to populations of DNA and RNA molecules. Nevertheless, systems for directing the evolution of proteins and other types of macromolecules are already on the horizon.

Directed evolution offers a way for the biochemist to interact with nature on its own terms. New macromolecules that evolved in the laboratory can bind specifically to macromolecules that evolved in nature. As natural macromolecules continue to change, directed evolution should permit the creations of researchers to keep pace—a point that could be critical in stalemating the natural evolution of drug resistance in viruses and other pathogens.

For example, through directed evolution, a biochemist could develop an RNA that binds to a viral protein and blocks infection by the virus. The viral protein could mutate into a resistant form, rendering the RNA drug useless. Yet directed evolution could generate new RNAs against the mutated form. This game of evolutionary cat and mouse can continue indefinitely.

An important future application of directed evolution technology will be the development of novel catalysts. Biochemists are already engaged in what has become known as "the rational design of enzymes": they carefully tinker with biological catalysts to alter their structural and functional properties. Yet if a large heterogeneous population of molecules and a practical selective amplification strategy is at hand, why not let the evolving molecules do the tinkering themselves? This twist places the biochemist in the unaccustomed role of spectator, carrying out what Brenner has wryly termed "the irrational design of enzymes."

It remains to be seen which catalytic functions will be accessible through directed evolution. The stunning accomplishments of Darwinian evolution in nature should be encouraging, but we should not forget that nature has had a four-billion-year head start. One catalytic function that is a high-priority target for investigation is the ability of a macromolecule to catalyze its own replication. Such molecules would begin to evolve in a self-sustained fashion—indeed, theorists generally agree that all life on the earth must have evolved from molecules that had this property. From a biochemical standpoint, such self-replicating molecules would be alive. It would be ironic if directed evolution, which began as an attempt to imitate life, turned out to be a way of reinventing it.

FURTHER READING

AN EXTRACELLULAR DARWINIAN EXPERIMENT WITH A SELF-DUPLICATING NUCLEIC ACID MOLECULE. Donald R. Mills, Roger L. Peterson and Sol Spiegelman in *Proceedings of the National Academy of Sciences,* Vol. 58, No. 1, pages 217–224; July 15, 1967.

SYSTEMATIC EVOLUTION OF LIGANDS BY EXPONENTIAL ENRICHMENT: RNA LIGANDS TO BACTERIOPHAGE T4 DNA POLYMERASE. Craig Tuerk and Larry Gold in *Science,* Vol. 249, pages 505–510; August 3, 1990.

SELECTION OF SINGLE-STRANDED DNA MOLECULES THAT BIND AND INHIBIT HUMAN THROMBIN. Louis C. Bock, Linda C. Griffin, John A. Latham, Eric Vermaas and John J. Toole in *Nature,* Vol. 355, No. 6360, pages 564–566; February 6, 1992.

IN VITRO GENETICS. Jack W. Szostak in *Trends in Biochemical Sciences,* Vol. 17, No. 3, pages 89–93; March 1992.

DIRECTED EVOLUTION OF AN RNA ENZYME. Amber A. Beaudry and Gerald F. Joyce in *Science,* Vol. 257, pages 635–641; July 31, 1992.

What Columbus "Saw" in 1492

When Christopher Columbus arrived in the New World, he brought with him a rich lode of cultural preconceptions that strongly influenced his perceptions of the land and its inhabitants

by I. Bernard Cohen

This year, the 500th anniversary of Christopher Columbus's famous voyage, has been the occasion for much reflection on the true nature of the admiral's achievement. Many writers have remarked that the "New World" found by Columbus was actually an old world that had long been inhabited by a culturally diverse, native civilization. It was also an old world in a less obvious sense: Columbus's perceptions of the lands he discovered were profoundly influenced by his intellectual and cultural preconceptions. In that way, his reactions were like those of any person confronting the unfamiliar, whether an explorer seeking new realms of the earth or a scientist trying to fathom the mysteries of nature. Ideas derived from the Bible, from the reports of previous adventurers, from mapmakers and from general lore all worked their way into Columbus's "discoveries."

Much of the extant information about Columbus's thoughts regarding the New World is contained in two documents, a diary and a letter. Unlike most mariners of the time, Columbus kept a regular day-to-day record of his 1492 voyage; he may in fact have been the first seafarer to keep such a log. On his return to Spain, Columbus presented the log to his patroness, Queen Isabella, who had a copy of it made for Columbus and retained the original for herself. Although both versions have disappeared, Bartolomé de las Casas, a historian and missionary known as the Apostle of the Indies, made his own version of Columbus's text (part transcription, part summary), copies of which still survive. Las Casas, who was the first priest to be ordained in the New World, devoted his later life to the cause of Native Americans and preached against their enslavement.

Early in 1493, shortly after his return to Spain, Columbus wrote a lengthy letter, a popular and summary version of his log. This letter was addressed to Luis de Santángel, who had helped Columbus solicit funds for his voyage, and to several others. The letter was widely read and was printed a number of times. It became the primary source of information about the first discoveries made by Columbus.

Modern readers of the two documents may be surprised by Columbus's general lack of interest in details concerning the lands he had visited and by the scant attention he paid to the animals and plants there. His descriptions of locations are so laconic that additional research is often required to determine exactly where Columbus was at the time of writing. For instance, on October 12, 1492, after a 36-day journey from the Canary Islands, Columbus landed on an island

I. BERNARD COHEN is Victor S. Thomas Professor (emeritus) of the History of Science at Harvard University. He has recently completed a new translation of Newton's *Principia*, to be published by the University of California Press; he has also edited a volume of critical and historical studies on the interactions between the natural sciences and the social sciences. This article on Columbus is based on Cohen's long-term study of creativity in science.

NATIVE AMERICANS were initially drawn based on ancient expectations and on sketchy information relayed by Columbus. This single-leaf illustration, published in Augsburg in 1505, shows the natives as cannibalistic primitives. The accompanying text relates that they have no government or private property.

he called Guanahaní (now known as San Salvador), one of the Bahama Islands. His main description is remarkably brief: "This island is very large and very flat and with many green trees and much water and a very large lake in the middle, without any mountain; and everything [so] green that it is a pleasure to look at it."

When Columbus did pay attention to details of geography, he generally noted only the outstanding features of the land. He also indulged in the kind of exaggeration that is the accompaniment of wonder. He wrote that Guanahaní has "a port large enough for as many ships as there are in Christendom," an overstatement comparable to Marco Polo's estimate that the shipping on the Yangtze River exceeded that on all the European waterways put together.

Columbus's observations were colored by his search for useful goods to bring back to King Ferdinand and Queen Isabella, who had financed his trip. His reports therefore include mentions of gold, silver, pearls and gems. But Columbus did not recognize the greater significance of the islands he had discovered. He was seeking the kingdom of the Great Khan, the islands of Japan and the riches of India. What he found was not what he wanted to discover, so the details of the land were, in that sense, irrelevant. In contrast, he had come across a strange new people who were not Europeans or Africans, yet neither were they Asians. The Native Americans attracted Columbus's attention all the more so because their natural beauty contradicted the possibility that he had reached a forbidden part of the earth inhabited by monsters. For these reasons, Columbus recorded in vivid detail the appearance of the natives he encountered without evincing too much concern for the land they inhabited.

Columbus's earlier reading reinforced his belief that his ocean voyage had brought him most of the way to the Far East. He made extensive marginal notes in his copy of *Imago Mundi* of Pierre d'Ailly, a 15th-century geographer. In

chapter 11, Columbus writes (partly quoting and partly paraphrasing the text): "The limit of the habitable earth toward the east and the limit of the habitable earth toward the west are quite close, and in between is a small sea." This passage encouraged Columbus in his plan to sail westward from Europe to China or to the region of India.

A disregard for geographic details was common among other explorers of Columbus's time. Amerigo Vespucci, for instance, wrote exquisitely detailed accounts of the natives of South America but gave only cursory information about the land itself. Not until well into the 16th century, when the Americas came to be recognized as previously unknown continents, did explorers begin to pay full attention to the flora and fauna of the New World.

In marked contrast to his lack of concern regarding the natural aspects of the islands he visited, Columbus showed acute interest in his encounters with the people who came to be mistakenly known as Indians. As Leonardo Olschki, the Italian historian of science, observed, Columbus was "meticulous and exhaustive" in reporting the appearance of the natives, their customs and peculiarities, "even depicting their life and habits with a keen and expressive realism."

Columbus's perceptions of the Native Americans were guided by the tales of his seafaring predecessors, by Judeo-Christian mythology and by his own expectations. The books that survive from the library of Columbus include annotated copies of works by Pliny, Aeneas Sylvius (Pope Pius II), Pierre d'Ailly and Marco Polo. Polo repeatedly told of strange sights in strange lands. Nevertheless, readers of his *Milione* cannot help but be struck by his constant comparison of what he witnessed on his travels with similar beings or objects at home. He thus sought to render the unfamiliar acceptable to the mind by relating bizarre features and exotic experiences to the ordinary life of the writer and the prospective readers. Columbus behaved in much the same way.

Based on his intellectual and cultural background, Columbus might have anticipated meeting five distinct kinds of people in the course of his 1492 journey. First, if he had indeed reached the Far East, as he expected and desired, the natives would have had to be Asians. The first natives he met, the Tainu on San Salvador, were obviously not the highly civilized denizens of India, China or Japan. Columbus tried to console himself by searching for hints that he might have reached islands off the coast of Asia.

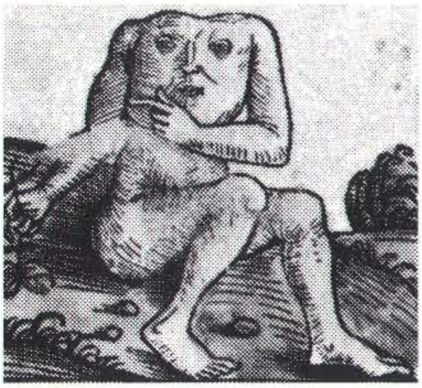
Second, the natives whom Columbus encountered might have been men and women of some other familiar type (Europeans, perhaps, or Africans), in which case he would have reached not the Indies but some unknown pocket of the world. Columbus's log clearly dispels such a thought. Equally unpleasant was the third, opposite possibility that he would confront a previously unrecognized race of men and women—that is, he might have reached some terra incognita. Columbus's unwillingness to admit that potentiality is well documented in his log.

A fourth prospect was that the newly discovered peoples would be inhabitants of an earthly paradise. One of the most enduring images associated with the Bible is the Garden of Eden, where a human couple, sublime in their nakedness, inhabited an idyllic wilderness. Columbus may well have been thinking of that image when he commented repeatedly on the surprising nakedness of the native people he confronted. Subsequent writers went much further in drawing parallels between Eden and the New World.

The final possibility that Columbus faced was that he might have reached one of the distant, forbidden parts of the world where monsters dwell. Tales of humanoid monsters were a common component of the travel literature and legends with which Columbus was familiar. They became an important part of Columbus's writings as well.

The mythical monsters known to Columbus included giants, one-eyed cyclopes and hairy men and women, as well as more exotic creatures. Amazon fighting women cut off the right

HUMANOID MONSTERS had been associated with remote parts of the world since at least the time of ancient Greece. The 1493 edition of Hartmann Schedel's *Nuremberg Chronicle* recorded many such monsters, including sciopods (*top left*), cynocephali (*middle left*), blemmyae (*bottom left*), cyclopes (*bottom center*) and panotii (*bottom right*).



MAP OF THE NEW WORLD made by the Turkish cartographer Piri Re'is in 1513 (*left*) incorporates information from Columbus's charts of America. Re'is identified South America as the home to monsters such as cynocephali (*detail, top right*) and blemmyae (*detail, bottom right*). He did so in defiance of an Islamic injunction against creating images of living beings.

breast so they could use bows and arrows more efficiently. Anthropophagi devoured human flesh and drank from human skulls. Blemmyae had heads located in their chests. Panotii were endowed with gigantic ears that they used as blankets or as wings for flying. Cynocephali had human bodies but dogs' heads. Sciopods possessed a single leg and a huge foot; they would stretch on their backs and hold the foot above themselves to act as parasols.

Information about those beings appeared in many written accounts and literary works. The alleged correspondence of Alexander the Great, for example, and Pliny the Elder's *Natural History* contain early mentions of humanoid monsters. Marco Polo's *Milione* included descriptions of monstrous races. Anyone who had read about distant places could have expected that such creatures would dwell there. References to monsters were still very much in evidence in 15th-century scholarly books such as d'Ailly's *Imago Mundi* and the *Historia Rerum Ubique Gestarum* of Aeneas Sylvius, both of which were read by Columbus.

Medieval maps gave graphic expression to the ideas concerning the strange beings that appeared in the literature. The earth of the medieval cartographers included terrifying boundaries of haunted seas inhabited by loathsome creatures. The Hereford map of the world, made in the late 13th century, depicted sciopods, blemmyae and other monsters dwelling in far-off lands. Other maps, both earlier and later, offered similar visual representations. Monsters were both a source of fear and of expectations for sailors.

Columbus transcribed and summarized the information about monsters that he found in *Imago Mundi,* indicating the passages that he found most significant. In a postil to chapter 12, Columbus writes that "at these two extremities [toward the north and south] there are savages who eat human flesh; they have vile and horrible faces. The cause is the intemperate climate; because of this they have bad habits and are savage." And in his marginalia to chapter 16, entitled "The Wonders of India," Columbus notes what he might expect to see in that land: pygmies, tall Macrobians, barbarians who kill and eat their elderly relatives, women who give birth only once and produce offspring with white hair that grows dark as they age, as well as the river Ganges, near which d'Ailly locates people who live on the smell of a certain fruit.

When Columbus reached the New World, he inquired again and again about the presence of humanoid monsters. Perhaps his informants did not understand what he was asking, or perhaps they attempted to please him by telling him what he seemingly wanted to hear. Columbus recounts in his letter that he received information about people with tails, people having no hair and women living and fighting on an island devoid of men, for example.

The tales to which Columbus paid attention, and the manner in which he in-

AMAZON WARRIORS were anticipated by many explorers of the New World. In his *Universal Cosmography* of 1575, André Thevet illustrates the barbaric nature of such native warriors in an imaginary battle scene set on the American mainland. Thevet writes that, unlike other races, the American Indians never make peace.

terpreted them, undoubtedly reflected both his expectations and his hopes. The poor communication between Columbus and the Indians (often based on just a few words and on sign language) gave him considerable leeway in imposing his own meanings on the Indians' stories. His wavering attitudes toward those stories expressed both his need to consider seriously the existence of monsters and his desire, for various practical reasons, to deny their existence.

While in Cuba, on November 4, one of the Indians who had been taken on board described a place where "there were one-eyed men, and others, with snouts of dogs, who ate men." Columbus later noted that the captive Indians feared the people living on Bohío (Hispaniola) "who had one eye in their foreheads, and others whom they called cannibals." The word "cannibal" was thus introduced into Western language. Columbus judged his informants were lying; he believed the people whom they dreaded and who had captured some of them "must have been under the rule of the Great Khan." His desire to believe he was near Asia outweighed his inclination to accept the reality of the monsters. In this instance, Columbus manifested a fascinating mix of what he actually heard, what he expected to find, and what he feared to encounter.

In his letter to Santángel, Columbus felt compelled to discuss the reports of monsters, if only to dismiss them as mere rumor. "In these islands I have so far found no human monstrosities, as many expected, but on the contrary the whole population is very well-formed," he wrote. He added that "as I have found no monsters, so I have had no report of any, except in an island Quaris, which is inhabited by people who are regarded in all the islands as very fierce and who eat human flesh." To prove to Ferdinand and Isabella, and to the people of Europe in general, that he had not traveled to lands inhabited by monsters, Columbus brought home to Spain some captive Indians who could display their well-formed bodies.

Columbus also specified in his log that some of the inhabitants of the islands he visited were quite submissive. I cannot help but take note of this aspect of Columbus's character. Of course he lived in an age when there was as yet no general belief in human freedom and dignity. But Columbus did more than acquiesce in the standards of his age; his first thought on encountering a group of simple people was of enslaving them. Columbus referred again to the docility of the Tainu. "They bear no arms, and are all unprotected," he recorded for the information of the king and queen, and are "so very cowardly that a thousand could not face three [armed Spaniards]." He concluded that "they are fit to be ordered about and made to work, to sow and do aught else that may be needed."

Such sentiments distressed Las Casas, who later wrote that "the natural, simple and kind gentleness and humble condition of the Indians, and want of protection, gave the Spaniards the insolence to hold them of little account, and to impose on them the hardest tasks they could, and to become glutted with oppression and destruction ... the Admiral enlarged himself in speech more than he should, and what he here conceived and set forth from his lips was the beginning of the ill usage he afterwards inflicted upon them." Clearly, the ability to lead a successful expedition does not necessarily imply nobility of character. Or, as Sigmund Freud commented of Columbus, "great discoverers are not necessarily great men."

Most of the time, however, Columbus's observations of the natives were connected to his effort to understand what part of the world he had reached. He described in detail the physical characteristics of the local inhabitants in order to prove that they were not deformed and that they differed in appearance from the known European or African peoples. After his first meeting with the natives of Guanahaní, on October 12, Columbus again wrote that they "are very well-formed, with handsome bodies and very fine faces." He noted that their "hair is not kinky but straight, and coarse like horsehair." They "wear their hair short over their eyebrows, but they have a long hank in the back that they never cut."

Columbus remarked that all the people of Guanahaní were "tall people and their legs, with no exceptions, are quite straight"; none was seen to have a paunch. Their eyes, he found, "are large and very pretty." He did judge, however, that "their appearance is marred somewhat by very broad heads and foreheads, more so than I have ever seen in any other race." He later learned that this feature of their appearance came from the pressure of a board on the forehead of infants. Columbus also recorded that some of the natives had painted their faces, others their whole bodies.

During the same leg of his journey, Columbus observed that the natives' "skin is the color of Canary Islanders or of sunburned peasants, not at all black." This last attribute, like many of the others he had entered in his log, struck him as important in establish-

ing that the island on which he had landed was not a part of Africa. He then added, "Nor should anything else be expected since this island is on an east-west line with the island of Hierro [Ferro] in the Canaries," a reference to Aristotle's theory that all forms of life at the same latitude should be identical.

Some days later, when Columbus arrived in Cuba, he observed that "the people are meek and shy" and "go naked like the others"; he found them to be "without weapons and without government." He may well have been making an implicit comparison to the Garden of Eden. Just before his departure from Cuba, on November 6, Columbus made a final entry about the natives. "These people are free from evil and war," he wrote, adding that they are "not as dark as the people of the Canaries." Some observations were discreetly omitted from Columbus's log. Samuel Eliot Morison, one of the foremost biographers of Columbus, pointed out that Columbus refrained from mentioning "any sporting between the seamen and the Indian girls," whose habits were "completely promiscuous," no doubt because "his Journal was intended for the eyes of a modest Queen."

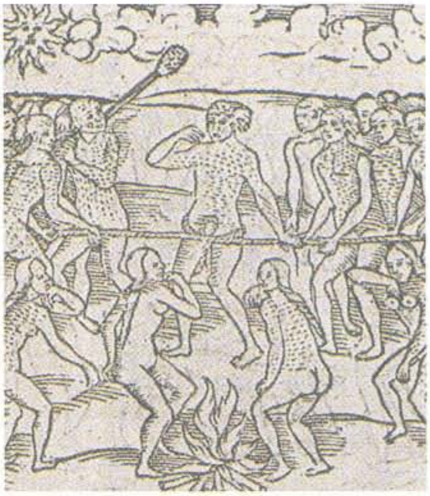
Despite Columbus's numerous statements in his log and letter about the true nature of the natives he encountered, other writers continued to endow the New World with the kind of monsters that for centuries people had come to regard as inhabiting such remote lands. In 1493 Giuliano Dati made a verse adaptation of Columbus's letter, a very popular work that rapidly went through four editions. Dati then wrote two poems that may be considered sequels, the second of which described monsters supposedly found in the East Indies. The second book even presented woodcuts depicting eight of the humanoid creatures.

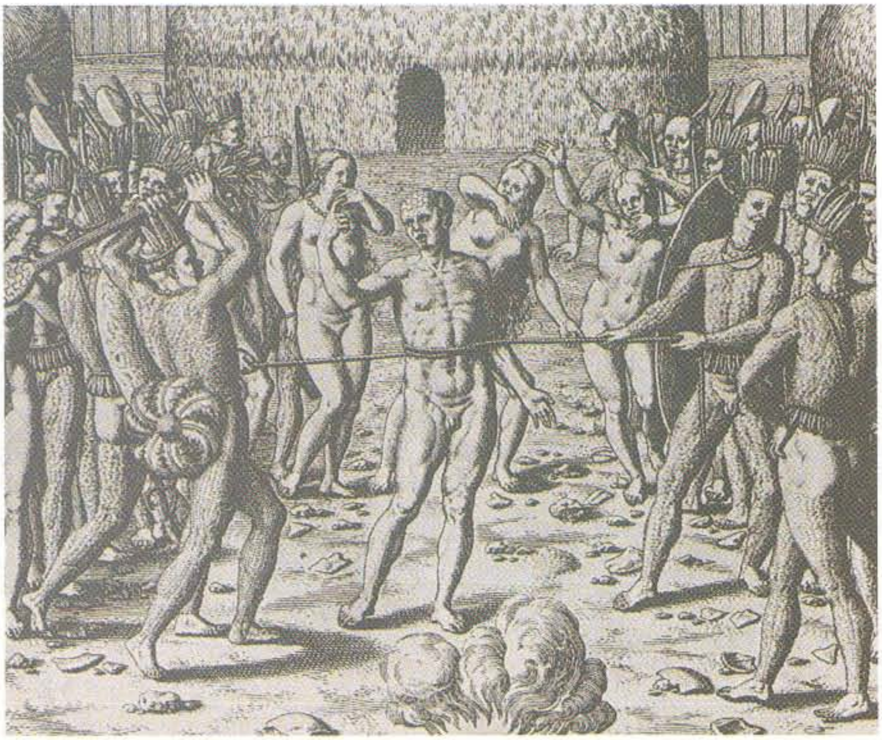
The force of popular tradition also is seen in a famous map of the Americas made in 1513 by Piri Re'is, a Turkish cartographer who wrote about the discoveries of Columbus. The map delineated the east coast of South America with an impressive degree of accuracy. On the land, the map portrays various real creatures such as parrots and monkeys. But he also included a unicorn, a dog-headed man and even a version of a man with a head beneath his shoulders.

Other explorers continued to entertain serious thoughts that monsters might reside in the New World. In 1518 Diego Velásquez de Cuéllar, the governor of Cuba, specifically told Hernán Cortés to watch for people with gigantic ears and dogs' faces. In 1522 Cortés sent back to King Charles V huge fossil bones supposed to have come from the skeletons of giants. The Amazon River was given its name by Francisco de Orellana, who believed it flowed through a region of warrior women lacking breasts and therefore resembling those described by ancient chroniclers. I suspect he mistook male Indian warriors wearing elaborate headdresses for women and therefore for Amazons.

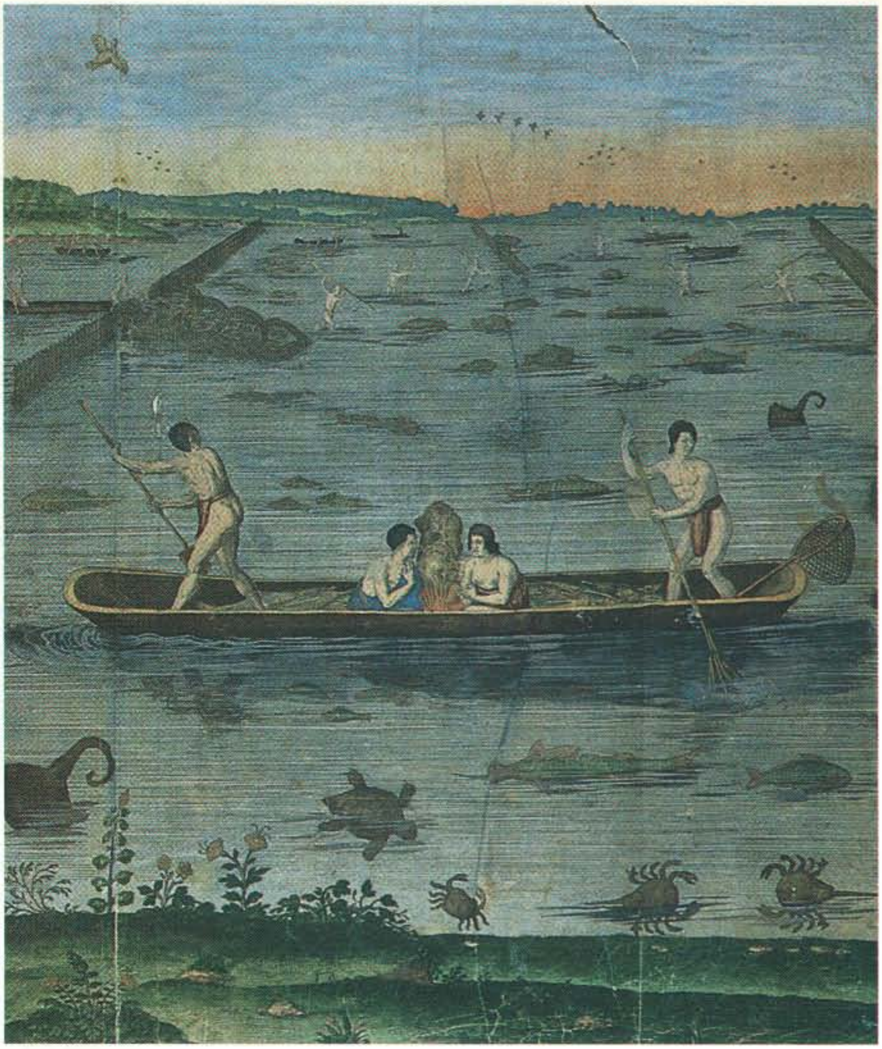
Although Columbus did not find monsters, he did bring back news of living beings even more unexpected: the "Indians," who were unlike any known human race in their appearance and customs. The reaction of European society illustrates the human tendency to search for familiar elements in unfamiliar information; in the words of physicist J. Robert Oppenheimer, "We cannot, coming into something new, deal with it except on the basis of the familiar and the old-fashioned."

The Bible offered one source of inspiration for understanding the Native Americans. Letter-writers of the time commenting on Columbus's expedition frequently cited the nakedness of the Indians, sometimes adding details about the use of a leaf or a swath of cloth to hide the genitals. Morison remarked that the "one touch of nature that made all newsmongers kin was the naked natives, especially the women who wore nothing but a leaf." He observed that the "points in Columbus's discovery that chiefly interested people were the new things that recalled something very old, like Adam and Eve in the Garden of Eden." Morison noted that "the lack of religion among the natives, their timid and generous nature, and ignorance of lethal weapons" were characteristics that, "combined with their prelapsarian innocence, suggested to anyone with a classical education that the Golden Age still existed in far-off corners of the





CANNIBALISM figured prominently in Western tales of remote parts of the globe. Hans Staden's 1557 woodcut (*above*) illustrates his alleged experiences in the Americas. In 1592 Theodore de Bry reworked Staden's drawing into a more refined image (*right*); in the process, he gave the figures an incongruous, classical European look.



SYMPATHETIC PORTRAYALS of life in America began to appear late in the 16th century. Theodore de Bry based this 1590 illustration on a watercolor of native Virginians by John White. In addition to capturing details of the natives' peaceful everyday existence, de Bry evinces a level of interest in the local flora and fauna absent from earlier depictions of the New World.

globe." He thus stated explicitly a theme implicit in Columbus's log.

Columbus's insistence that the people of the New World were perfectly formed, beautiful human beings had an interesting by-product. Artistic representations of Native Americans made by the next generation of explorers highlighted that perfection of form, transforming it into a positive statement rather than a mere denial. A result is that 16th- and 17th-century artists depicted Native Americans as if each one were a Greek athlete posing for the sculptor Phidias. In this sense, the earthly paradise had indeed been found.

The nature of kinship between Europeans—that is, Spaniards—and the people of the New World became a central issue in the debate over how to introduce the Native Americans to Christianity. In 1550 Las Casas argued that the Indians were fellow humans and that they should accordingly be treated with tolerance and kindness, leading up to their peaceful conversion to Christianity. Juan Ginés de Sepúlveda, canon of Salamanca and royal historian of Spain, responded that even if those creatures were "more men than beasts"—of which he was by no means certain—they were without question of an inferior sort. Their natural condition was servitude toward their Spanish conquerors, and their conversion should be effected by force, he believed.

The superior virtue of the Spaniards made them fit to rule over the Indians, Sepúlveda continued, just as was the case for the Greeks ruling over the barbarians. He characterized the Indians as showing "every kind of intemperance and wicked lust," including cannibalism. Among the Spanish virtues were "prudence, talent, magnanimity, temperance, humanity and religion"—qualities that seem rather ironic, especially given the history of the treatment of the Native Americans by the Spanish conquistadors.

Attitudes toward the culture of the Indians also varied considerably. The 16th-century Spanish explorers José de Acosta and Gonzalo Fernández de Oviedo y Valdés attempted to record, in a relatively sympathetic manner, the beliefs, practices and social organization of the Native Americans. At the same time, Bishop Diego de Landa of Mexico considered some of the habits and customs of the Mayans so savage and repulsive that he systematically destroyed documents recording their civilization.

These divergent interpretations of the people of the New World should come as no surprise. A succession of philosophers of science—most recently Thomas S. Kuhn—have convincingly demonstrated that even the ostensibly objective facts of science are to a large degree laden with theory. The history of science shows that the force of theory in interpreting experiments and observations is much like the power of preconception so abundantly evident in Columbus's accounts of the natives of the New World and in those produced by his contemporaries.

Columbus hoped to discover the realm of the Great Khan, and to his death he fervently believed he had done so. He knew that he might encounter monsters and heard stories that exactly fitted his concerns. Even as he began to learn something about the beliefs and customs of the Native Americans, Columbus imposed his expectations and pragmatic goals on his perceptions of them. In a real sense, the root of his achievement may have resided in the way he combined his fears and his beliefs so that, both in spite of and because of them, he not only committed grave errors but also achieved marvelous discoveries.

FURTHER READING

ADMIRAL OF THE OCEAN SEA: A LIFE OF CHRISTOPHER COLUMBUS. Samuel Eliot Morison. Little, Brown, 1942.

THE MONSTROUS RACES IN MEDIEVAL ART AND THOUGHT. John Block Friedman. Harvard University Press, 1981.

THE LOG OF CHRISTOPHER COLUMBUS. Translated by Robert H. Fuson. International Marine Publishing, 1987.

THE FOUR VOYAGES OF COLUMBUS. Translated and edited by Cecil Jane. Dover Publications, 1988.

MARVELOUS POSSESSIONS: THE WONDER OF THE NEW WORLD. Stephen Greenblatt. University of Chicago Press, 1991.

The Human Voice

How the voice works was largely unknown until modern technology became available. New instruments are now improving the care and treatment of the voice

by Robert T. Sataloff

Twenty years ago the human voice was a mystery. Little was known about how it works or how to care for it, despite centuries of fascination with the voices of singers and actors and the crucial importance of vocal communication to society. Literature on voice medicine, and particularly on the care of the professional voice, was scarce. In the few papers that did exist, there was scant emphasis on how the voice worked.

The state of therapy was equally weak. Nonsurgical treatments of benign vocal-fold problems were controversial, and the surgery available involved vocal-cord stripping. (Specialists replaced the term "cord" about 10 years ago with the more descriptive term "fold.") This procedure ripped away the superficial layers of the vocal fold under the assumption that healthy tissue would grow to replace the unhealthy tissue. Unfortunately, many patients ended up permanently hoarse, although their vocal folds afterward looked normal.

Since that time, a new medical subspecialty has emerged. Spurred by interest in the problems of professional singers and actors, scientific and technological advances have raised the standard of care for all voice patients. These improvements were made possible by interdisciplinary collaborations among professionals, who, at first, barely spoke the same language. The Voice Foundation, which was established by physician Wilbur James Gould of New York City to promote such interactions, held its first symposium in 1972 and brought together laryngologists, voice scientists, speech pathologists, singing and acting teachers and performers. The exchange of ideas at that meeting led to new collaborations, new directions in research and many major advances.

ROBERT T. SATALOFF is professor of otolaryngology at Jefferson Medical College of Thomas Jefferson University in Philadelphia and editor of *The Journal of Voice.* He is also a professional singer and singing teacher and serves on the faculty of the Curtis Institute of Music and of the Academy of Vocal Arts. In addition, he is a university choir and orchestra conductor. He has published more than 150 scientific articles and 10 books, including *Professional Voice: The Science and Art of Clinical Care.*

Today, 20 years later, it is possible for a singer with a few "lost notes," a governor running for president, a salesperson whose voice is weak, a smoker with a tumor or anyone else with a voice complaint to get sophisticated medical attention. That care is a result of the growing understanding of how the voice works.

The vocal mechanism involves the coordinated action of many muscles, organs and other structures in the abdomen, chest, throat and head. Indeed, virtually the entire body influences the sound of the voice either directly or indirectly. For grasping the vulnerabilities of the vocal tract, a brief tour of this complex, delicate mechanism is necessary. The first stop, and the best-known part of the mechanism, is the larynx, or voice box.

The larynx has four basic anatomic components: a cartilaginous skeleton, intrinsic muscles, extrinsic muscles and a mucosa, or soft lining. The most important parts of the laryngeal skeleton are the thyroid cartilage, the cricoid cartilage and the two arytenoid cartilages. The extrinsic muscles connect these cartilages to other throat structures; the intrinsic muscles run between the cartilages themselves.

One pair of intrinsic muscles extends from the arytenoid cartilages to a point inside the thyroid cartilage, just below and behind the Adam's apple. These thyroarytenoid muscles form the bodies of the vocal folds; the space between them is the glottis. The vocal folds are normally the source of the human voice.

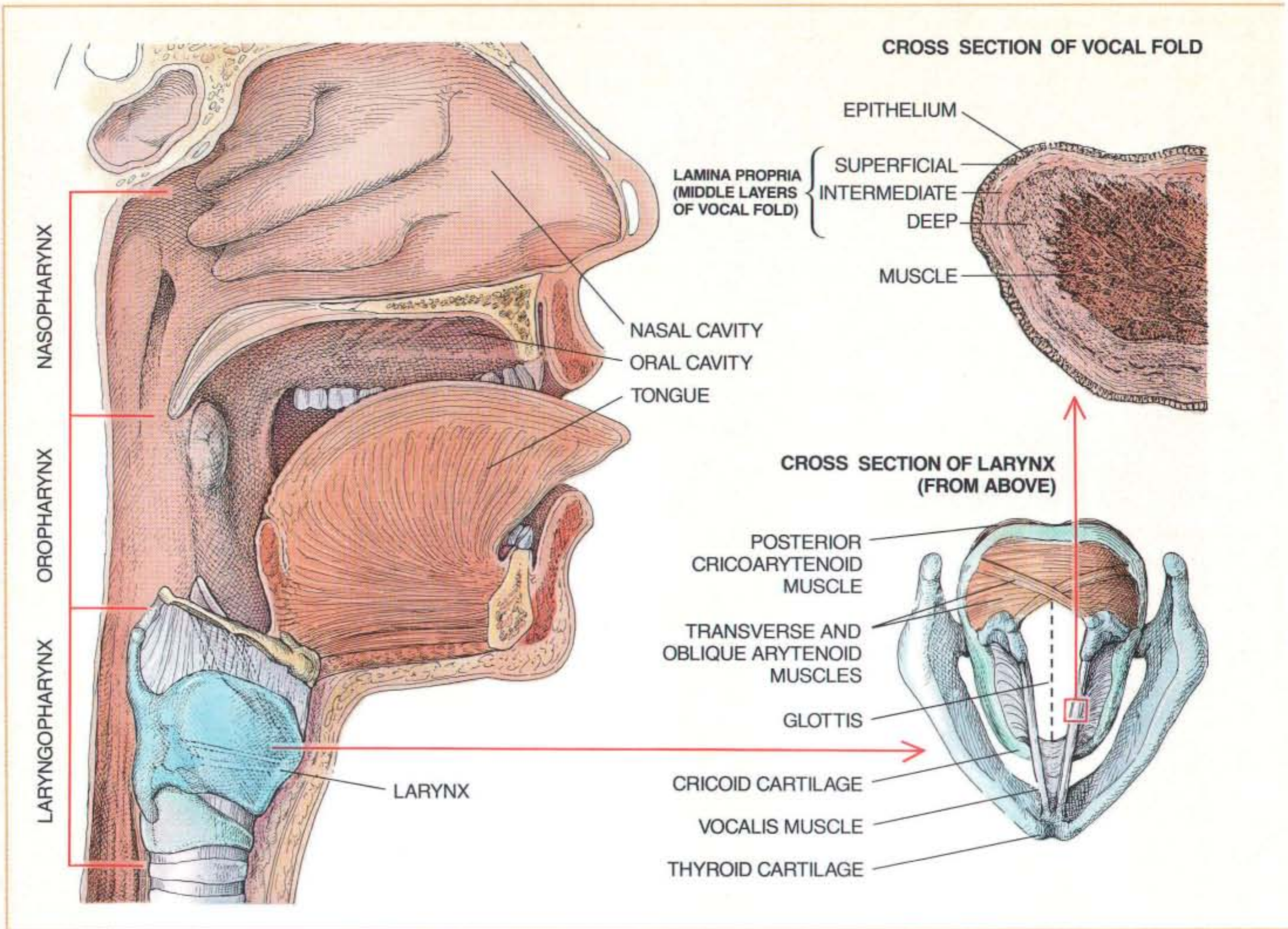
The intrinsic muscles can change the relative positions of the cartilages and pull them through a range of motions. These changes alter in turn the shape, position and tension of the suspended vocal folds. The cricothyroid muscle, for example, participates in the control of pitch by increasing the longitudinal tension (stretching) of the vocal folds.

The extrinsic muscles, also known as the strap muscles of the neck, raise and lower the laryngeal skeleton. The resulting accordion effect also changes the angles and distances between the cartilages and alters the resting lengths of the intrinsic muscles. The larynx has a natural tendency to rise and fall as the pitch of the voice ascends and descends. Such large adjustments in position, however, interfere with the fine control over the vocal folds that is essential for smooth vocal quality. For that reason, classically trained singers are generally taught to use their extrinsic muscles to maintain the laryngeal skeleton at a fairly constant height regardless of pitch. This technique enhances a unified vocal quality throughout a singer's range.

The soft tissues lining the larynx are much more complex than had been thought. The mucosa forms the thin, lubricated surface of the vocal folds that makes contact when they are closed. The mucosa overlying the vocal folds is different from that lining the rest of the larynx and respiratory tract: it is stratified squamous epithelium, which is better suited to withstand the trauma of vocal-fold contact.

The vocal fold is not a simple muscle covered with mucosa. In 1975 physician Minoru Hirano of Kurume, Japan, identified five distinct tissue layers in the

VOCAL-FOLD SURGERY prolonged the career of the popular singer Elton John. He had trouble with his voice during a tour of the U.S. in 1986. The problem turned out to be a nonmalignant lesion, which surgeons successfully removed early in 1987. A year later he was able to resume giving concerts.



structure. Beneath the thin, lubricated epithelium on the surface lie the superficial, intermediate and deep layers of tissue called the lamina propria. Underlying the lamina propria is the thyroarytenoid (or vocalis) muscle itself. The five layers have different mechanical properties that produce the smooth shearing motions essential to healthy vocal-fold vibrations.

When the vocal folds vibrate, they produce only a buzzing sound. That sound resonates, however, throughout the supraglottic vocal tract, which includes the pharynx, the tongue, the palate, the oral cavity and the nose. That added resonance produces much of the perceived character and timbre, or vocal quality, of all sounds in speech and song.

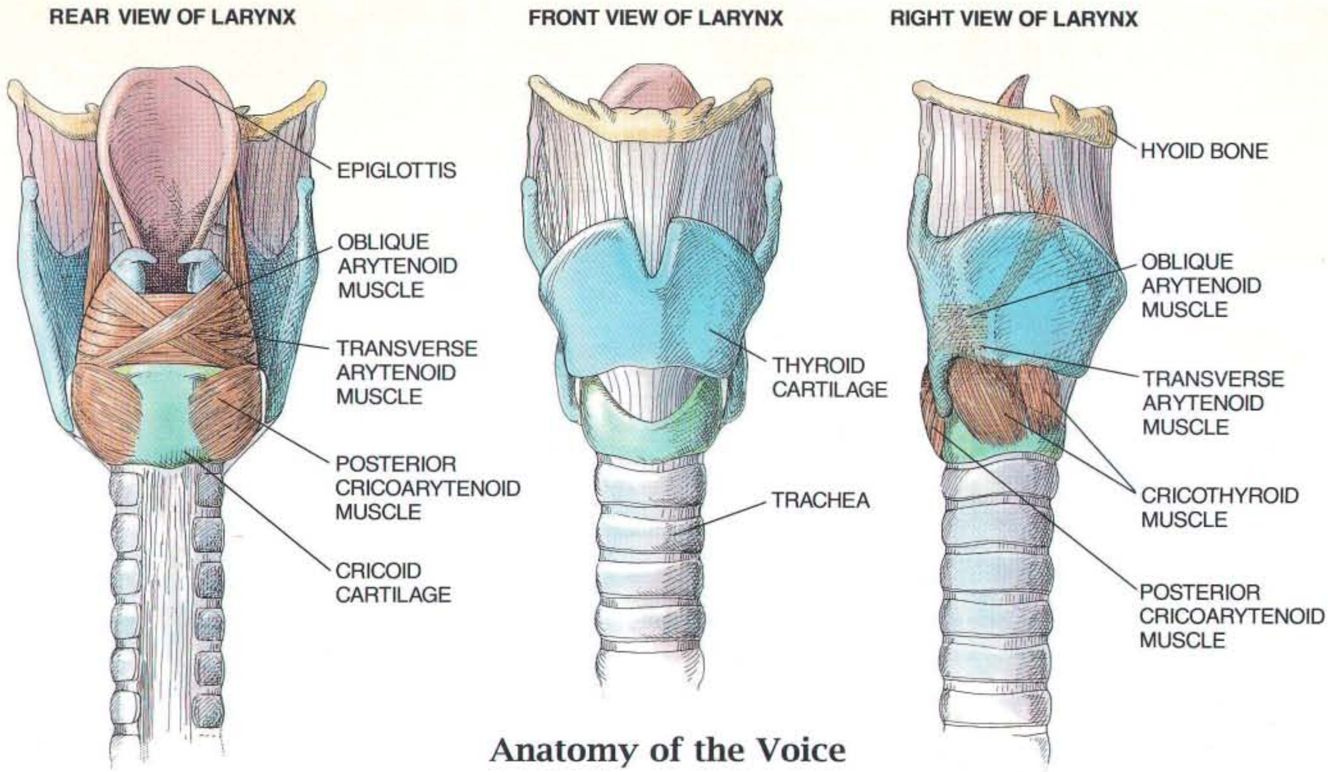
The power source for the voice is the infraglottic vocal tract—the lungs, rib cage and abdominal, back and chest muscles that generate and direct a controlled airstream between the vocal folds. As the glottis closes, opens and alters shape, its air resistance changes almost continuously. The power source must therefore make rapid, complex adjustments to maintain a steady vocal quality. Singers and actors refer generally to the entire power complex as their "support" or "diaphragm." Actually, the anatomy of the power complex is complicated and not completely understood, and performers who use such terms do not always mean the same thing.

The principal muscles of inspiration, or inhalation, are the diaphragm (a dome-shaped muscle that extends along the bottom of the rib cage) and the external intercostal (rib) muscles. Expiration, or exhalation, is largely passive during quiet respiration: the mechanical properties of the lungs and rib cage typically force air out of the lungs effortlessly after a full breath. Of course, active expiration is also possible, and many of the muscles involved in this process are also used to support voice production, or phonation.

During active expiration, muscles may raise the pressure within the abdomen and thereby force the diaphragm upward. Alternatively, they may lower the ribs and sternum to decrease the dimensions of the thorax. The primary muscles of expiration are the abdominal muscles, but internal intercostals and other chest and back muscles also contribute.

Trauma or surgery that alters the structure or function of these muscles undermines the power source of the voice, as do asthma and other diseases that impair expiration. People often compensate for deficiencies in their support mechanism by overworking their laryngeal muscles, which are not designed to serve as a vocal power source. Such behavior can result in decreased function, rapid fatigue, pain and even structural problems, such as vocal-fold nodules.

Like the muscular and skeletal systems, the nervous system also contributes to voice production. The "idea" for a voice sound originates in the cerebral cortex and travels to motor nuclei in the brain stem and spinal cord. These areas send out complicated messages for coordinating the activities of the larynx, the thoracic and abdominal musculature and the vocal-tract articulators. Signals from certain divisions in the nervous system, called the extrapyramidal



Anatomy of the Voice

Vocal mechanism encompasses many muscles and organs of the abdomen, chest, throat and head. The drawing at the far left portrays those in the throat and head. Details of the larynx, or voice box, are shown to the right of that drawing in an orientation looking down on the structure with the front—the Adam's apple—facing the bottom. The two vocalis muscles constitute the bodies of the vocal folds, which were formerly known as the vocal cords; a cross section of one of them appears above the representation of the larynx. The remaining three drawings (*above*) show the major muscles and cartilages of the larynx from the rear (*left*), the front and the right side.

tract and the autonomic nervous system, also refine these activities.

The nerves that control the muscles of the vocal tract are potential sources of voice problems. For example, the two recurrent laryngeal nerves control most of the intrinsic muscles in the larynx. Because those nerves (especially on the left) run through the neck, down into the chest and then back up to the larynx, they are easily injured by trauma or surgery on the neck and chest.

Nerves also provide feedback to the brain about voice production. Auditory feedback, which is transmitted from the ear through the brain stem to the cerebral cortex, allows a vocalist to match the sound produced with the sound intended. Tactile feedback from the throat and muscles also may help with the fine-tuning of vocal output, although that process is not fully understood. Trained singers and speakers cultivate their ability to use tactile feedback effectively because they expect that poor room acoustics, loud musical instruments or crowd noises will interfere with the auditory feedback.

During phonation, all those anatomic structures and systems must work together. The physiology of voice production is exceedingly complex, but the voice can be likened to a trumpet. Power for the sound is generated by the chest, abdomen and back musculature, which produce a high-pressure airstream. A trumpeter's lips open and close against the mouthpiece to create a buzz similar to that produced by the vocal folds. This sound then resonates through the rest of the trumpet, which is analogous to the supraglottic vocal tract.

Much of the progress during the past 20 years has come from filling in the details of how vocal sounds originate and change. Part of this effort has involved modeling the movements of the vocal folds. Although a vocal fold has a five-layer anatomy, it behaves mechanically more like a three-layer structure, consisting of a cover (epithelium and superficial layer of the lamina propria), a transition layer (intermediate and deep layers of the lamina propria) and a body (thyroarytenoid muscle).

Observations and modeling studies have revealed how the larynx produces a sound. Initially, the vocal folds are in contact, and the glottis is closed. As the lungs expel air, pressure below the glottis builds, typically to a level of about seven centimeters of water for conversational speech. This pressure progressively pushes the vocal folds apart from the bottom up, until the glottis is open and air begins to flow. Elastic and other forces resist the separation of the upper margin of the vocal folds, but the airstream overpowers them.

The flow of air, however, produces a Bernoulli effect—that is, a reduction in the lateral air pressure caused by its forward motion. The effect tends to pull the vocal folds shut, as do the elastic properties of the vocal-fold tissues. The pressure of the airstream below the glottis also diminishes as the glottis opens to let the air out.

Because of these factors, the lower edges of the vocal folds begin to close almost immediately, even though the upper edges are still separating. That closure further diminishes the force of the airstream. The upper margins of the vocal folds then snap back to the mid-

line and close the glottis. Subglottal pressure builds again, and the events repeat themselves. (It should be understood that there is direct pressure and that the alternating variations rarely drop the subglottal pressure to zero. This fact is important in understanding the driving forces involved in the motions of the vocal fold.)

An important aspect of this process is that the lower part of the vocal folds begins to open and close before the upper part. The rippling displacement of the vocal-fold cover produces a wave motion in the mucosal layer. If the complex vibration of that glottal wave is impaired, hoarseness or other changes in voice quality may result.

The vocal folds do not excite the air by vibrating like violin strings. Instead, by opening and closing the glottis, they create puffs of air in the vocal tract. The sudden cessation of airflow at the end of each puff produces an acoustic vibration. The mechanism is similar to that which causes the sound of hand clapping.

The sound from the larynx is a complex tone containing a fundamental frequency, or pitch, and many overtones, or higher harmonic partials. (Frequency is measured in hertz, the number of opening and closing cycles in the glottis each second.) Surprising as it may seem, trained and untrained vocalists produce about the same spectrums at the vocal-fold level.

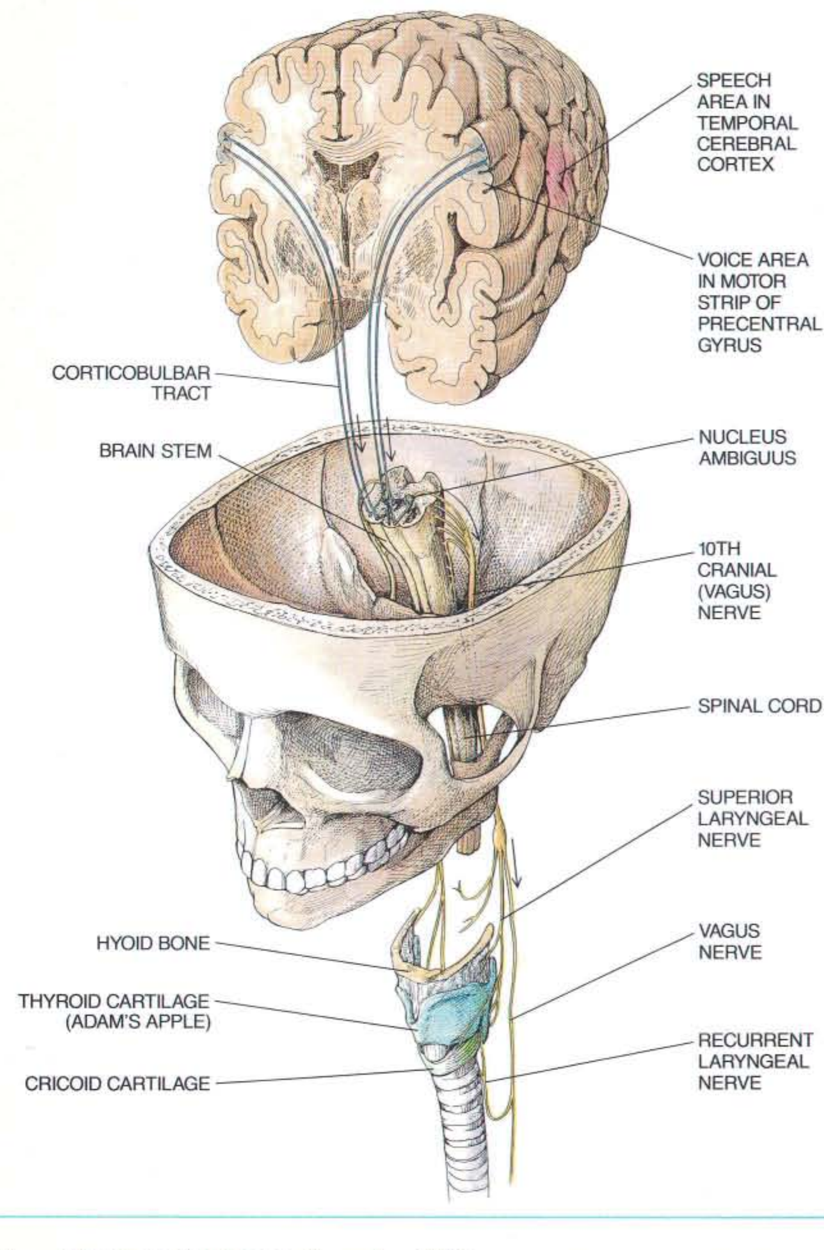
The pharynx (the throat area between the mouth and the esophagus), the oral cavity and the nasal cavity act as a series of interconnected resonators for the voice signal. The system is more complex than a trumpet because its walls, and hence its shape, are flexible. In any resonator, some frequencies are attenuated while others are enhanced, or radiated with higher amplitudes. Certain harmonic partials therefore become relatively softer while others grow louder. Johan Sundberg of the Royal Institute of Technology in Stockholm has shown for singers (and his colleague Gunnar Fant for speakers) that the vocal tract has four or five important resonance frequencies called formants. The intensity of the voice source diminishes uniformly across its frequency spectrum except at the formant frequencies, where it peaks.

Formant frequencies are established by the shape of the vocal tract, which can be altered by the laryngeal, pharyngeal and oral cavity musculature. Overall, the length and shape of one's vocal tract are individually fixed and determined by age and sex: women and children have shorter vocal tracts than do men and consequently have higher formant frequencies. Nevertheless, the dimensions of the vocal tract can be consciously adjusted to some degree, and mastering those adjustments is fundamental to voice training.

One resonant frequency that has received attention is known as the singer's formant. It appears to be responsible for the "ring" in a singer's or speaker's voice. The ability to make oneself heard clearly even over an orchestra depends primarily on the presence of the singer's formant: there is little or no

How the Voice Is Produced

The production of speech or song, or even just a vocal sound, entails a complex orchestration of mental and physical actions. The idea for making a sound originates in the cerebral cortex of the brain—for example, in the speech area. The movement of the larynx is controlled from the voice area and is transmitted to the larynx by various nerves. As a result, the vocal folds vibrate, generating a buzzing sound. It is the resonation of that sound throughout the area of the vocal tract above the glottis—an area that includes the pharynx, tongue, palate, oral cavity and nose—that gives the sound the qualities perceived by a listener. Auditory feedback and tactile feedback enable the speaker or singer to achieve fine-tuning of the vocal output.



significant difference in the maximum vocal intensities of trained and untrained singers.

The singer's formant occurs at around 2,300 to 3,200 hertz for all vowel sounds. Aside from adding clarity and projection to a voice, it also contributes to differences in timbre. The singer's formant occurs in basses at about 2,400 hertz, in baritones at 2,600 hertz, in tenors at 2,800 hertz, in mezzo-sopranos at 2,900 hertz and in high sopranos at 3,200 hertz. The singer's formant is often much less pronounced in sopranos.

Control over two vocal characteristics—fundamental frequency and intensity—is crucial. One way to raise the fundamental frequency is to raise the pressure of the airstream moving through the larynx. For mechanical reasons, as the pressure rises, the vocal folds tend to blow apart and to snap shut more quickly and frequently. Singers learn to compensate for this tendency: otherwise, their pitch would rise whenever they tried to sing louder.

Generally, the most efficient technique for altering the pitch is to change the mechanical properties of the vocal folds. Contracting the cricothyroid muscle makes the thyroid and cricoid cartilages pivot and stretches the vocal folds. This change exposes more surface area on the vocal folds to the airstream and thereby makes them more responsive to air pressure. It also stretches the elastic fibers of the vocal folds and increases their efficiency at snapping back together. The pitch rises because the cycles of opening and closing in the glottis (phonatory cycles) shorten and repeat more frequently.

Vocal intensity, or loudness, depends on how much the vocal-fold vibrations excite the air in the vocal tract. Raising the air pressure increases the amplitude of the vibrations because the vocal folds move farther apart and snap together more briskly. Consequently, during each phonatory cycle, the flow of air through the larynx cuts off more sharply, and the intensity of the produced sound increases. A similar effect increases the intensity of the sound of hand clapping.

A useful biophysical indicator of the efficiency of vocal control strategies can be seen in flow patterns during each phonatory cycle. For example, a vocalist may attempt to increase vocal intensity by excessively raising both the air pressure and the resistance of the glottis to the flow of air, using the muscles of the infraglottic vocal tract and the adductory (glottis-closing) forces of the vocal folds. Such a combination of forces results in a condition called pressed phonation, in which the amplitude of the voice's fundamental frequency is low despite considerable physical effort.

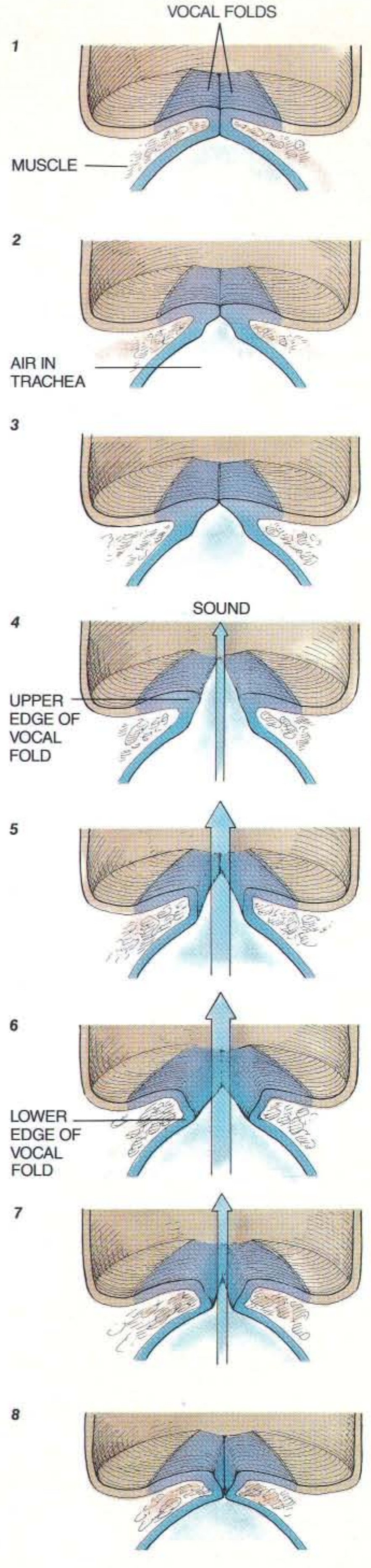
The amplitude of the voice source will also be low if the adductory forces are so weak that the vocal folds do not make contact and the glottis becomes inefficient. This condition results in breathy phonation. In contrast, a third and more desirable condition known as flow phonation is characterized by low airstream pressure and low adductory force, which increase the intensity of the fundamental frequency and make the voice louder. To identify pressed, breathy or flow phonation, voice specialists can plot changes in the flow of air through the glottis, thus producing a graph called a flow glottogram.

Sundberg has shown that a vocalist can raise the amplitude of the fundamental frequency by 15 decibels or more simply by changing from pressed phonation to flow phonation. Hence, people who rely habitually on pressed phonation expend unnecessary effort to achieve a loud voice. The forces and patterns of muscle use recruited to compensate for this inefficiency may damage the larynx.

By understanding the vocal control mechanisms, physicians can detect and correct the problems that abuse the voice and traumatize the vocal folds. Understanding the function of each component of the vocal tract also aids the development of optimal strategies for rehabilitating damaged voices.

The development of new tools has been critical for the science of the voice. Until the 1980s, the physician's ear was routinely the sole instrument used to assess voice quality and function. Practical techniques for observing and quantifying voice functions were generally lacking.

In 1854, for instance, a singing teacher named Manuel García devised the technique of indirect laryngoscopy. He used the sun as a light source and a dental mirror placed in a student's mouth to look at the vocal folds. Indirect laryngoscopy rapidly became a basic tool for physicians, and it is still in



VIBRATION OF VOCAL FOLDS is shown, in a vertical cross section through the middle part of the vocal folds, during the production of a single sound. The perspective is from the front of the larynx. Before the process starts (*1*), the folds are together. They separate as air is forced upward through the trachea (*2–7*) and then come together again as the sound ceases (*8*).

daily use. (Today, of course, we use an electric light instead of the sun.)

Yet as valuable as this technique has been, it has many disadvantages. Effective magnification of the vocal folds and photographic documentation of their condition are difficult. Also, standard lighting does not permit physicians to see the rapid, complex vibrations of the vocal folds.

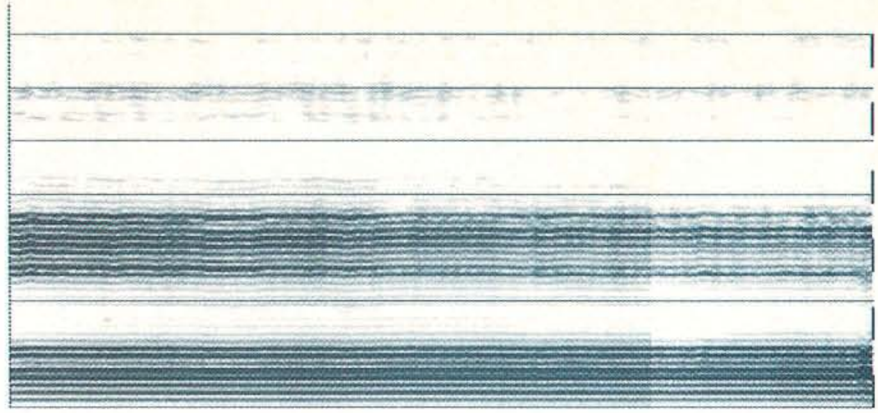
Currently the primary technique for inspecting vocal-fold vibrations is strobovideolaryngoscopy. It uses a microphone placed near the larynx to trigger a stroboscope that illuminates the vocal folds. If the frequency of the stroboscopic light is about two hertz out of phase with the vibration, an observer can watch the vocal folds in simulated slow motion. A crude version of this technique was actually first developed in the 19th century. Only during the past decade, however, have stroboscopes become bright enough and video cameras sufficiently sensitive for it to be useful routinely.

The stroboscopic effect permits detailed evaluation of the vocal-fold edge. Physicians can see small masses, vibratory asymmetries, scars, early carcinomas and other laryngeal abnormalities—many of which are not detectable under normal light. Digital analysis of the images can also supplement visual assessments, although poor image resolution and some other problems have limited the value of the technique so far.

Another method for monitoring vocal-fold vibrations is electroglottography. A weak high-frequency voltage between two electrodes placed on the neck passes through the larynx. Changes in the measured voltage generate a wave on the electroglottograph that illustrates vocal-fold contact. Information about the open glottis can be inferred from photoglottography, which measures light passed from below the vocal folds, or from flow glottography.

Measurements of aerodynamic function, which include comprehensive testing of pulmonary function and laryngeal airflow, are especially valuable. Together they reveal both the function of the vocal power source and the efficiency of the vocal folds for controlling airflow. Measurements of phonatory ability—the ability to produce sounds—are simple and useful for quantifying vocal dysfunctions and evaluating the results of treatment. Such tests determine the frequency and intensity ranges of the voice, how long a sound can be produced and other factors.

Laryngeal electromyography, another technique for studying voice function,



HEALTHY AND AILING VOICES are compared in these sonograms made as the speakers produced the sound of "a" as in "father." Time runs from left to right for about two seconds. The horizontal lines mark off frequencies in hertz from zero at the bottom to 7,000 at the top. The sonogram at the left is from a male speaker

involves the insertion of thin electrodes into the laryngeal muscles. It is useful in specialized circumstances for assessing neuromuscular integrity and function. For instance, measurements of electrical activity in the laryngeal muscles may foretell a patient's recovery from vocal-fold paralysis. In that case, before considering surgery, a physician might recommend waiting for a spontaneous recovery.

A skilled laryngologist or another trained listener can glean much from the sound of a voice. Nevertheless, clinicians and researchers need equipment capable of quantifying the vocal characteristics that are meaningful to the ear. The available equipment is helpful, but it needs further improvement.

For example, acoustic spectrography displays the frequency and harmonic spectrum of a voice and visually records noise. The equipment depicts the acoustic signal and enables researchers or physicians to make generalizations about the vocal quality, pitch and loudness. A variety of qualities can be measured: the formant structure and strength of the voice, the fundamental frequency, breathiness, the harmonics-to-noise ratio (or clarity of the voice) and perturbations of the cycle-to-cycle amplitude ("shimmer") and of the cycle-to-cycle frequency ("jitter"). Subtle characteristics, however, still cannot be detected. For instance, in studies of voice fatigue in trained singers, the difference between a rested and tired voice is usually obvious to the ear, but significant changes cannot be detected consistently even with sophisticated equipment.

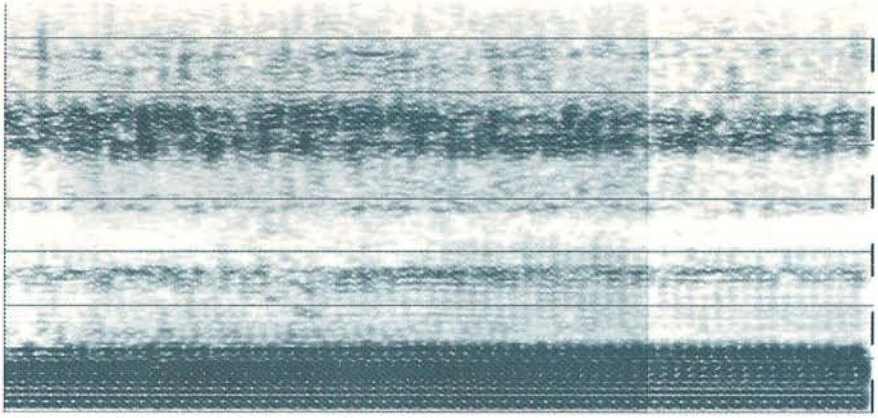
Psychological influences on the voice are also critical, but the techniques for evaluating them are poorly standardized. Nevertheless, well-developed questionnaires, tape recordings and assessment of voice by several observers have boosted the utility of such examinations. All these tools help physicians to detect and record the information contained in the sound of the voice more reliably and validly.

As technology has enhanced the diagnostic and therapeutic aspects of voice medicine, the need for laryngeal surgery has diminished. Some conditions require little more than the prescription of drugs. Medications must be used with caution, however, because even many over-the-counter remedies have side effects that adversely affect voice function. Antihistamines, for example, cause dryness in the vocal mucosa, which can lead to hoarseness and coughing. The anticoagulant properties of aspirin can contribute to vocal-fold hemorrhages.

Techniques have been developed to rehabilitate voices that have been damaged through misuse. Voice therapy facilitates breathing and abdominal support and helps to eliminate unnecessary muscle strain in the larynx and neck. It can even cure some structural problems of the vocal folds, most notably nodules (hard, callouslike growths). Therapy helps patients to learn how to use each component of the vocal tract appropriately so as to avoid straining and abusing their voices, how to maintain the correct levels of moisture and mucus in their vocal tracts and how to mitigate the effects of smoke and other hazards in the environment.

Good vocal hygiene and technique, however, are not always enough. Some structural problems in the larynx must be treated with surgery. These problems include nodules that have not responded to voice therapy, polyps (soft-tissue growths), cysts (fluid-filled masses) and other growths.

Most benign pathological conditions



with normal vocal folds; the one at the right is from a male speaker with growths on his vocal folds. His voice makes additional noise in the range above 5,000 hertz and has disrupted and weakened harmonics between 2,000 and 4,000 hertz. The result is that his voice sounds harsh and hoarse.

are conveniently superficial. By a variety of delicate microsurgical techniques, surgeons can now usually remove lesions from the epithelium or the superficial layer of the lamina propria without disrupting the intermediate or deep layers of the tissue, which form scars. Such procedures are now commonly described as phonosurgery. (That term originally referred only to operations designed to alter vocal pitch and quality.)

Most voice surgery is performed through the mouth while the surgeon views the larynx through a metal tube called a rigid laryngoscope. An operating microscope magnifies the larynx. A surgeon can then treat the laryngeal abnormalities with microscopic scissors, lasers and other instruments.

Nodules, polyps and cysts on the vibratory margin of the vocal folds are removed most safely with traditional surgical instruments. The operations can be remarkably precise: in some cases, it is possible to raise the vocal-fold mucosa, remove a cyst or other underlying mass and then replace the mucosa. Such minimally traumatic surgery does not even require postoperative voice rest, and rapid healing with good voice quality usually follows.

Lasers are often celebrated as revolutionary high-tech surgical instruments, but they are not always the best choice for laryngeal surgery. At the power densities required for the surgical ablation of tissue, the beam from a standard carbon dioxide laser would be surrounded by a heat halo as much as 0.5 millimeter wide. If the beam were directed against a lesion on the edge of the vocal fold, the heat might provoke scarring in the intermediate or deep layers of the lamina propria. Such a scar would create a nonvibrating segment on the vocal fold; hoarseness would result.

Nevertheless, the carbon dioxide laser is extremely useful for some lesions. It can seal off varicose blood vessels in the vocal folds that might hemorrhage, and it can vaporize blood vessels that nourish laryngeal polyps, papillomas and some cancers. Further research and development in laser technology is likely to provide an instrument better suited for microsurgery on the vocal fold in the near future.

New surgical techniques for modifying the laryngeal skeleton have been pioneered by physician Nabuhiko Issiki of Kyoto. These have become extremely useful for treating vocal-fold paralysis, which is a common consequence of viral infections, surgery or cancer. Traditionally, surgeons have treated vocal-fold paralysis by injecting small volumes of Teflon into the affected vocal fold. The Teflon pushes the paralyzed fold toward the midline of the glottis and allows the normal fold to meet it. The glottis can then close, and the patient's voice is often improved.

Yet although Teflon is relatively inert, tissue reactions to it are not uncommon. The stiffness that the Teflon produces in the vocal-fold edge frequently impairs the quality of the voice. Also, if the results of the Teflon injection are unsatisfactory, it is difficult to remove the material from the tissues.

For these reasons, Teflon injections have generally been replaced by thyroplasty. In this technique, a surgeon cuts a small window in the laryngeal skeleton and pushes the thyroid cartilage and the laryngeal tissues inward. The depressed cartilage is then held in place with an inserted Silastic block. Such an operation pushes the vocal fold toward the midline without injecting a foreign body into the tissues, and it appears to be more reversible than Teflon injection. My colleagues and I have also recently introduced an injection technique that uses, instead of Teflon, a small amount of fat harvested from the patient's abdomen or arm. Like the Teflon procedure, this one is relatively simple, and it lacks the disadvantages of Teflon. It requires further study.

Surgery on the laryngeal skeleton can also modify a patient's pitch. A surgeon can raise the pitch by pivoting the thyroid and cricoid cartilages and closing the space between them. These changes lengthen and tense the vocal folds. Alternatively, a surgeon can cut vertical sections out of the thyroid cartilage to shorten the vocal folds, decrease their tension and thereby lower the pitch. The results of such surgery are not sufficiently predictable for singers and other professional voice users to elect them for purely aesthetic reasons. Nevertheless, they are valuable for treating certain voice abnormalities and for adjusting the vocal pitch of patients who have undergone sex-change surgery.

Because most of the gains in treating the human voice have involved collaborations among physicians, voice scientists, speech-language pathologists, teachers and singers, they have found their way into practical use in medical offices unusually quickly. Moreover, educating patients, singing and acting students, voice teachers and the general public about the importance of the voice and its maladies has had gratifying results. Education is often the best preventive medicine, and it has already decreased the prevalence of avoidable voice problems.

For medical progress to continue, we will need even more basic understanding of voice science, better clinical evaluation and quantification tools, and better surgical instruments, such as more effective surgical lasers. We can anticipate not only further clinical advances but also exciting applications in voice training and development.

FURTHER READING

CLINICAL MEASUREMENT OF SPEECH AND VOICE. R. J. Baken. College Hill Press, 1987.

THE SCIENCE OF THE SINGING VOICE. J. Sundberg. Northern Illinois University Press, 1987.

PROFESSIONAL VOICE: THE SCIENCE AND ART OF CLINICAL CARE. R. T. Sataloff. Raven Press, 1991.

THE SCIENCE OF MUSICAL SOUNDS. J. Sundberg. Academic Press, 1991.

THE PRINCIPLES OF VOICE PRODUCTION. I. R. Titze. Prentice-Hall (in press).

VOICE SURGERY. W. J. Gould, R. T. Sataloff and J. R. Spiegel. C. V. Mosby Company (in press).

Experimental Market Economics

Laboratory microcosms of the commercial world shed light on the principles that govern trading and suggest ways to make real markets more effective

by Vernon L. Smith and Arlington W. Williams

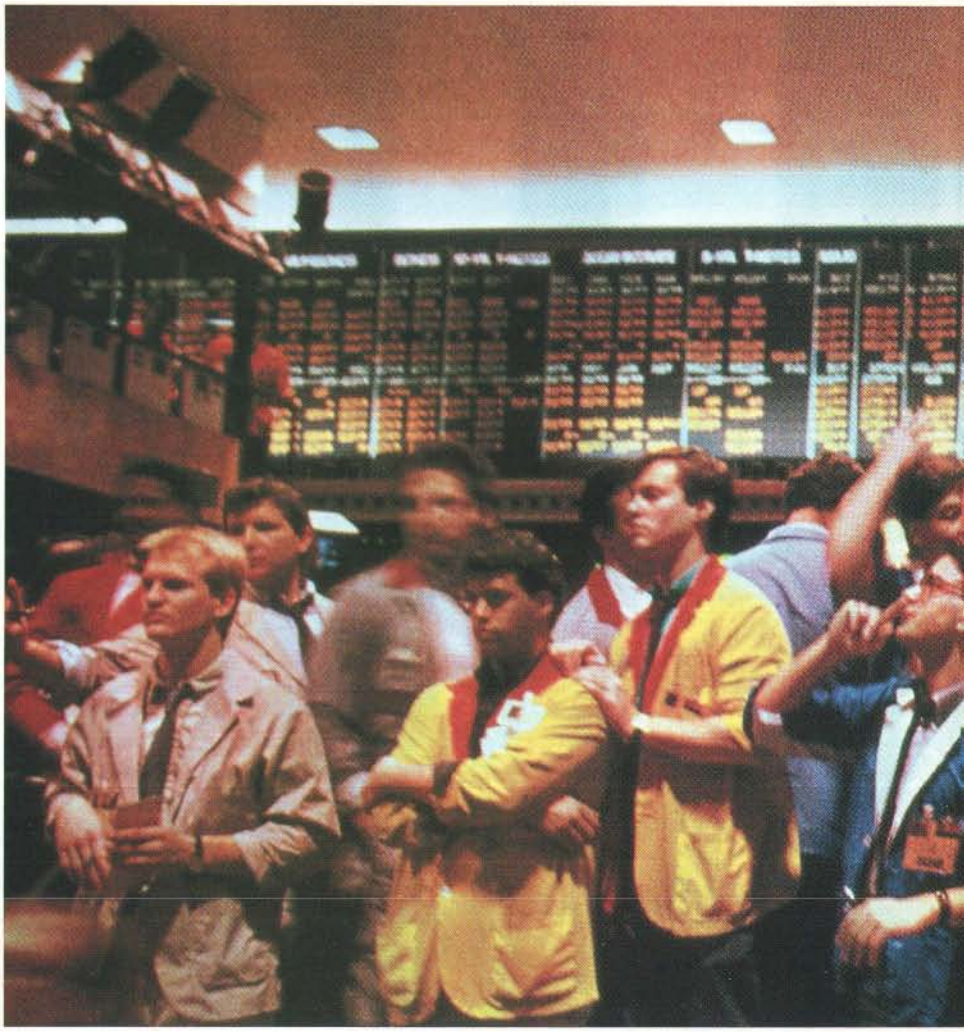
For thousands of years, people have been prone to "truck, barter and exchange," as Adam Smith wrote in 1776. Both parties to a voluntary, nonfraudulent transaction believe they are better off than before. Organized markets amplify the individual transaction and make it possible for people to trade what they produce for a wide range of goods and services produced by others. As a result, people can specialize in activities of their own choosing and still fulfill their needs for food, clothing, shelter and a host of necessities and luxuries. Specialization increases efficiency, and so the system of exchange creates wealth for society as a whole.

Central to this economic view of the world is the notion that markets set a price for goods that matches supply to demand. At the equilibrium price, producers are willing to sell precisely as much of a good as consumers are willing to buy at that price. Supply and demand, however, are not directly observable, and so for the most part this axiom has stood with virtually no empirical foundation. For most of its history, economics has differed fundamentally from the physical sciences and biology because economists did not conduct controlled laboratory experiments. Economists analyzing real markets cannot measure directly the functions that govern supply and demand; they can only infer them by observing prices and trading volume. As a result, their conclusions carry more than a hint of tautology.

That state of affairs has begun to change. During the past decade, market laboratories have been established at roughly a dozen universities and other institutions around the globe. In such facilities, economists can test the predictions of many of their theories. Workers using these laboratories can also explore ways for new technology to create more efficient markets.

Edward Chamberlin of Harvard University took the first steps toward experimental economics. In 1948 he reported what were probably the first market experiments. He set students to trading as a classroom exercise: the class was divided into "buyers" and "sellers" who paired up and attempted to negotiate a mutually agree-

VERNON L. SMITH and ARLINGTON W. WILLIAMS have collaborated in economic investigations for more than 15 years. Smith is Regents' Professor of Economics and research director of the Economic Science Laboratory at the University of Arizona, where he has taught since 1975. Williams, professor of economics at Indiana University, wrote his doctoral dissertation under Smith's supervision in 1978. He is involved in the design of computerized markets and has a particular interest in the use of laboratory experiments for economics education.



able price. The prices of successful transactions were posted, and unsuccessful traders split up to find new partners. Prices in these markets, however, did not converge to the theoretical equilibrium. Various explanations were offered for this discrepancy, among them the fact that the students had no knowledge of what the theoretical equilibrium price should be. In addition, unlike the traders in a real market, they had no monetary incentive to find it.

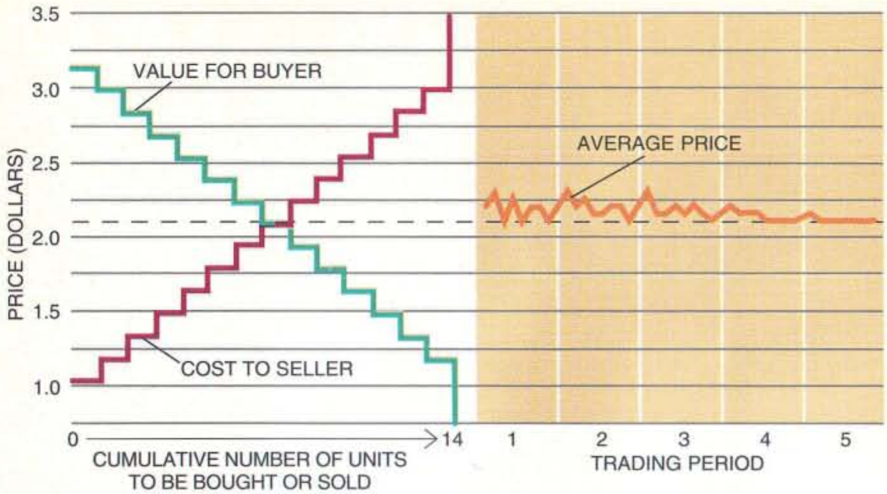
In 1956 one of us (Smith) set out to investigate Chamberlin's results further. One of Smith's experiments employed 14 buyers and 14 sellers who traded units of a hypothetical commodity. Each buyer received a card indicating the value of purchasing one unit, and each seller received a card indicating the cost for providing one unit. Buyers and sellers knew only the numbers on their own cards and had no idea what the equilibrium price might be. They were told they would receive a cash bonus at the end of the experiment equivalent to the difference between their cost (or value) and the price they negotiated in the open market. The values for the buyers ranged downward from \$3.15 to \$1.05 in increments of \$0.15, and the costs for the sellers ranged upward from \$1.05 to \$3.15 [*see illustration on next page*].

Whereas Chamberlin had found that the prices in his market fluctuated widely and never settled at the theoretical level, Smith found that his students' prices converged rapidly and reliably. The experimental conditions differed from Chamberlin's in a number of ways, but subsequent research has made it clear that the most important difference was the market institution by which price information is disseminated. Instead of pairwise negotiations, Smith set up a system in which any buyer or seller announced a bid or offer to the entire group, and a transaction occurred whenever any buyer accepted an offer or any seller accepted a bid. This system, called a double continuous auction, closely resembles the rules governing stock and commodity exchanges. It generates an action-packed, seemingly chaotic market and also makes a great deal of information available to all the participants. Knowing the bids and offers of the other traders helps each buyer or seller to bring the price closer to the competitive equilibrium.

Indeed, experimental market economists have found that the choice of institutions is often the essential factor in determining how a market works—whether trading proceeds smoothly and whether the market price does in fact converge to its theoretical level.

During the first decade of Smith's work, experimental markets consisted of a group of subjects in a room exchanging bids and offers while the experimenter tried valiantly to record all the information flowing by. In 1976,

TRADING FLOOR at the Chicago Mercantile Exchange appears chaotic, but the rapid flow of information among brokers helps to ensure that the price at which transactions take place matches supply to demand. The authors have found that the institutional rules governing trading play a crucial role in market efficiency.



SUPPLY AND DEMAND CURVES in an experimental market are set by giving each experimental subject a different internal cost or value for one unit of a hypothetical commodity (*left*). Sellers are willing to part with their unit at any price above their internal cost, and buyers are willing to purchase a share at any price below their internal value. At the theoretical equilibrium, supply and demand are equal. A record of trades in this experimental market (*right*) shows that trades do in fact take place at a price close to the equilibrium.

during the course of Williams's doctoral dissertation, we conducted the first computerized market experiments. By interposing the computer between buyers and sellers, we were able to record experimental data automatically. The computer records all information associated with the trades, even if it exceeds the experimenter's immediate needs. Thus, old experiments can be reanalyzed in light of new theories that make more subtle predictions.

Each computer terminal provides subjects with the appropriate information about the trading environment, such as a history of price movements or records of their own trades and assigned costs or values. In addition, the computer can present a series of instructional screens that subjects can read at their own pace, thus relieving the experimenter of having to explain how to trade and reducing any possible personal influence on the experiment.

The computer could enforce whatever set of market rules we chose, but we focused first on a computerized version of the double continuous auction. Even at the time, it was clear that the major financial markets would eventually evolve away from manual, face-to-face trading toward geographically dispersed computer networks. Mimicking their primary market institution, we hoped, would give us some insight into how such dispersed markets might behave.

Although computerized markets differ in many ways from their face-to-face predecessors, some characteristics persist. One phenomenon that has plagued organized trading almost since its inception is the speculative bubble: prices rise to unheard-of levels and then plunge without warning. In the 17th century it was tulip bulbs in Holland; on Black Monday in 1987 it was the U.S. stock market.

To make bubbles possible, we had to allow our subjects to attempt to buy low and sell high, thus realizing a capital gain. In earlier experiments, participants had remained in the role of buyer or seller throughout each session. Furthermore, sellers could not carry stocks over from one trading period to the next, making it impossible for them to speculate on possible price changes. Under the new rules, in contrast, traders could either buy, sell or hold units for future sale. As a result, the value they put on a unit—and the price they were willing to pay—depended on how much they believed others might be willing to pay for it later.

Over the centuries economists have advanced various theories of stock-market behavior. According to the prevailing "rational expectations" model, prices for an asset follow the return on investment that it produces. If all the traders in a market had the same expectation of a stock's return on investment, this theory predicts, bubbles would be impossible. To test this hypothesis, we designed a market in which participants could both buy and sell shares as well as hold them across several trading periods for their expected dividends or in anticipation of a rise in its price.

In a typical experiment, each "trader" received an initial, variable stake of cash and shares. At the end of each trading period, each share paid a dividend drawn from a probability distribution centered on a fixed value. The expected value of a single share's dividend stream (simply the average dividend times the number of trading periods remaining) was announced at the beginning of each trading period. For example, if the experiment was 15 trading periods long and the average dividend 24 cents, then everyone would be informed at the beginning of the first period that the "average holding value" of a share was \$3.60, at the beginning of the second period that the value was \$3.36, and so on. At the end of the experiment, each trader received his or her initial cash stake plus accrued dividends and capital gains, minus capital losses (in other words, the stock itself had no value at the end of trading).

Instead of controlling the cost and value of a share directly, this market experiment manipulated the share's underlying "intrinsic value," as determined by the expected stream of dividends. Traders, however, may base their decisions on expectations of capital gains as well as on intrinsic value, and so the supply and demand curves governing trades are no longer directly measurable.

According to the theory of rational expectations, all the traders have exactly the same information; consequently, expectations regarding future prices should be nearly homogeneous. If there is any trading at all, it should take place at prices very near the intrinsic value of a share, which falls linearly over the course of the experiment. As a result, no one will expect to make a capital gain from buying low and selling high, and little or no trading will occur. In contrast, if for any reason traders have diverse expectations about the intrinsic price, those who have low expectations will sell to those who have high ones.

The question we set out to answer was whether common information about the fundamental value of a share is sufficient to induce common expectations. It turns out that it is not. The standard rational-expectations model does very poorly at predicting the short-run price of a share when individual valuations depend strongly on expectations about the behavior of others. Although the rational expectations ultimately prevail, convergence is slow, and the lack of a theory to explain the sequence of prices satisfactorily is painfully evident.

We put subjects through the experiment three times on separate days. Each market cycle lasted between two and three hours. On the first day, trading started at prices well below intrinsic value, quickly rose to levels well above it and stayed there until the last

two trading periods, at which point the price crashed. The bubble recurred on the second day of trading, although its duration was considerably shorter than on the first day. Prices during the last few periods of the second trading session tracked the downward trend of intrinsic value. On the third day, prices at last roughly approximated intrinsic value from the beginning.

A key component of the rational-expectations prediction is that traders with the same information have similar beliefs about the future value of a share. We altered the experiment slightly, asking traders before each period to forecast the average share price, offering a small reward to the trader with the lowest cumulative error. This forecast may not match the traders' price expectations exactly, but it is the best measure we could devise. As expected, we found that inexperienced traders made widely diverse forecasts; after they had gone through two days of trading, they made predictions that converged on the shares' intrinsic value.

Curiously, subjects' perceptions of these markets provide little or no insight as to the cause of the speculative bubble. One subject asked, during the first trading period, "What is this buying panic? Shouldn't prices be around $3.60?"

Another said that his strategy was to buy below dividend value and sell above but that the market moved above dividend value before he could buy in. "So I bought anyway in hopes that it would rise higher."

"Prices rose without cause," a third subject said. "When the market turned down, I knew I had blown it, not selling earlier, but I just couldn't bring myself to sell even though prices were still above dividend value."

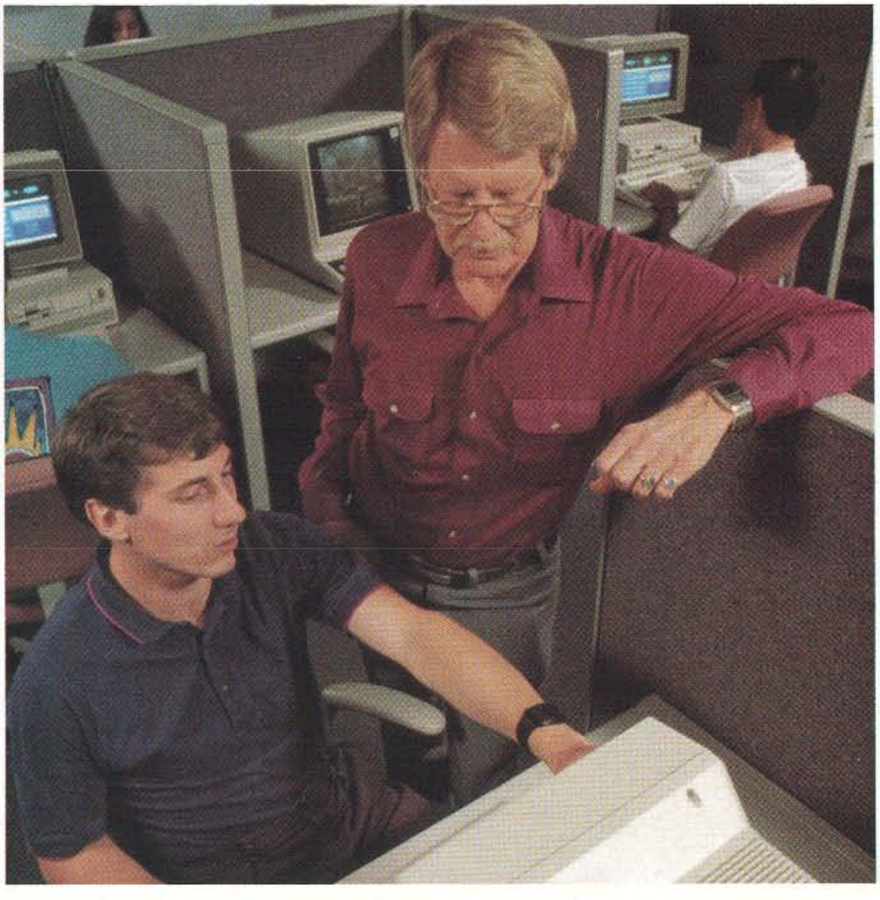
We have looked for ways to eliminate the bubble-crash phenomenon, but it appears to be very robust. Speculative bubbles occur in markets where traders can sell short (selling shares that they do not yet own in the anticipation that the price will drop by the time the shares must be delivered) and in markets where they can buy on margin (borrowing money against the value of currently held shares to purchase additional ones). Neither transaction fees for each trade nor rules that limit the maximum price change in any single period will squelch bubbles. The pattern of speculative rise followed by a crash persists when the market is converted to a form in which traders do not make public bids or offers. And perhaps most significant, bubbles appear whether the subjects of our experimental markets are students, business executives or professional stock traders.

No market institution appears able to eliminate speculative volatility in prices; nevertheless, the overall effectiveness of commonly used institutions does differ. We tested how well price levels and trading tracked changes in the theoretical equilibrium under three sets of rules. The first was the double continuous auction. The second was posted-offer pricing, in which sellers set a price and buyers decide how many units, if any, to purchase. The third was the double sealed auction, in which traders prepare bids and offers, and a third party executes the appropriate trades.

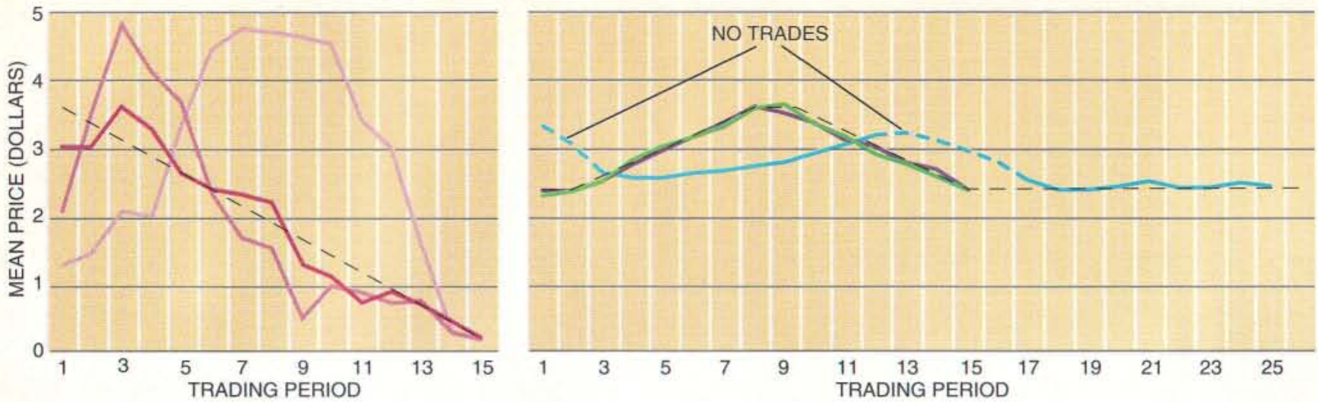
Our experimental market contained four buyers and four sellers, each of whom could trade as many as six units of a hypothetical commodity. The private valuations for each unit differed so that a trader would receive less profit buying or selling a second unit at the same price as the first, a third at the same price as the second, and so on. The private values that determined the market's supply curve were fixed, but those determining demand changed over the course of the experiment. After two periods of stability, demand increased for six trading periods, leveled off for two and then declined for another six periods.

When we set traders to work under the double continuous auction rules, prices during each trading period fluctuated significantly, but the prices of the final trades in each period were remarkably close to the competitive equilibrium [*see illustration on next page*]. Furthermore, the volume of trades also moved in the direction predicted by competitive price theory, increasing as prices rose and decreasing as they fell. The market was quite clearly in a state of disequilibrium, but its internal dynamics drove it quickly in the "proper" direction.

The double continuous auction is characteristic of stock and commodity markets. In contrast, most U.S. retail markets—from supermarkets to doctors' offices—use posted-offer pricing. The experiments we ran show that this institution does not track a changing equilibrium very well at all. Prices lagged while demand was rising, and yet at the same time, trading volume was low



EXPERIMENTAL SUBJECTS take part in trading at the University of Arizona's economics laboratory. Computers display information, transmit bids and offers among traders and then record all market activity for later analysis.



SPECULATIVE BUBBLE (*left*) occurs when investors bid the price of a commodity up above its intrinsic value (*black line*). As traders gained experience, the magnitude of the bubble decreased. No market institution appears capable of preventing bubbles, but institutions do differ in the speeds with which prices adjust to a changing equilibrium (*right*). Double continuous and sealed auctions cause prices to change quickly, whereas posted-offer pricing leads to market dislocations (*blue*).

compared with the number of trades that could have been profitably made. As demand declined, furthermore, prices continued rising until all the offers posted by sellers exceeded the value buyers put on the product. There were no trades during three consecutive periods, and posted prices declined rapidly. We allowed the market to continue for another 10 periods with constant demand until it finally stabilized. Prices at that point were slightly above the competitive equilibrium, and trading volume was slightly below it.

Even though posted-offer pricing was clearly inferior to the double continuous auction in our experiments, it may still be the institution of choice for large, stable markets—especially those in which each transaction involves only small sums. The dampening effect of posted-offer pricing on trading volume and its inability to track a shifting equilibrium are outweighed by the fact that it does not impose any negotiating costs. Stock and commodity markets, in contrast, impose high negotiating costs, employing thousands of people in full-time negotiations to maintain prices at their equilibrium level.

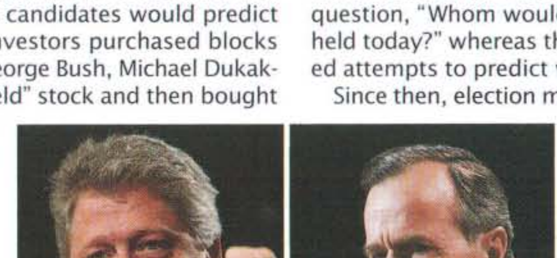
The last institution we studied was the double sealed auction. It appears to set prices as effectively as the double continuous auction, but its transaction costs are lower. Traders prepare a set of bids or offers specifying the number of shares they are willing to buy at any particular price; their personal workstations send these price-quantity combinations to the central

Putting Your Money Where Your Mouth Is

Experimental markets can also be useful for investigating aspects of human behavior not directly connected with economic theories. During the five months preceding the 1988 presidential election, a team of economists, political scientists and programmers at the University of Iowa created a presidential stock market to see how well the "market prices" for various candidates would predict the outcome of the election. Investors purchased blocks containing one share each of George Bush, Michael Dukakis, Jesse Jackson and "rest of field" stock and then bought or sold the individual shares in hopes of maximizing their gain. Each candidate's share paid a single postelection dividend of \$2.50 multiplied by his percentage of the popular vote; total dividend payments thus equaled the funds invested.

The market institution employed was a variant of the double continuous auction. Investors could make bids or offers from terminals connected to a university minicomputer or from their homes by means of a modem. The market predicted the outcome of the election more accurately than did standard opinion polls, and candidates' share prices fluctuated far less than their percentage ratings in the polls. Researchers suggested one reason for this result: polls reflect answers to the question, "Whom would you vote for if the election were held today?" whereas the market reflected profit-motivated attempts to predict who would actually win.

Since then, election markets have proved more accurate than polls in predicting a gubernatorial contest in Illinois and parliamentary elections in Germany. Nearly 1,000 people participated in the 1992 presidential market in which winning shares are worth \$1 and others nothing. As of October 16, shares of George Bush and Bill Clinton were trading at 14 cents and 83 cents, respectively, and late entrant H. Ross Perot was trading at 3 cents.



\$0.83 **\$0.14**

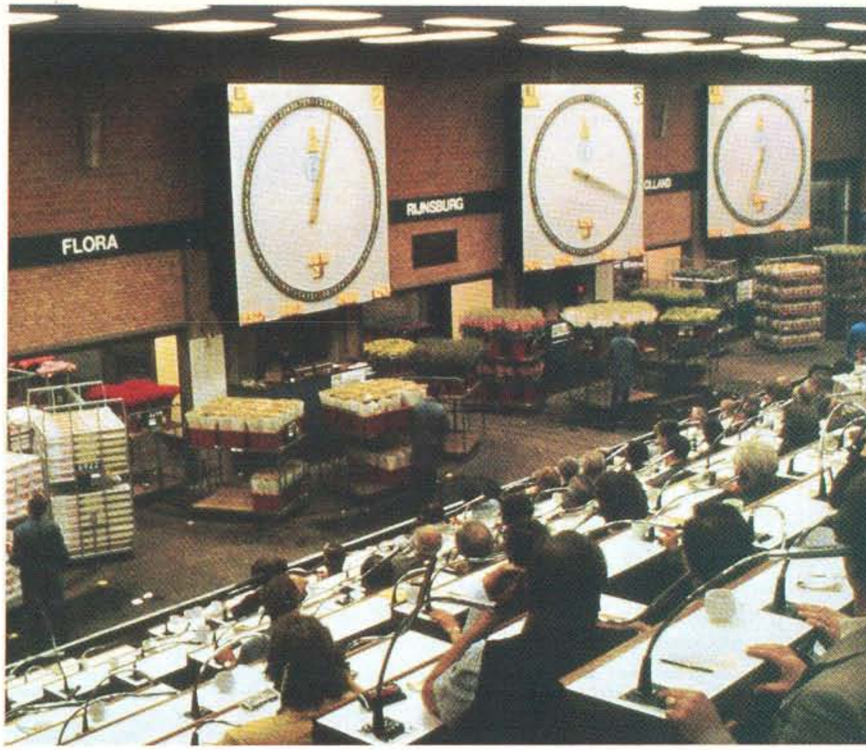
market computer at fixed intervals. The central computer sets the market price by finding the intersection of the curves marking the cumulative number of units for sale at any given price and the cumulative number of units desired. Orders to buy that specify a price above the intersection of the two curves are executed, as are orders to sell that specify a price below the market level. All trades take place at the market price—not the price specified in the buy and sell orders. Orders thus indicate the highest acceptable purchase price or lowest acceptable sale price. If there is an excess of supply or demand at the market price, the computer randomly eliminates orders to match the two.

After the trades have been completed, the central computer informs traders of the market price, the volume of shares traded and the status of their trades. If the highest-priced buy order is less than the lowest-priced sell order, then no shares change hands, and traders are informed of the high bid and low offer.

In our experiments the double sealed auction tracked the competitive-equilibrium price very closely. Indeed, the variance of the prices in the double sealed auction was less than in the double continuous auction. This finding suggests that stock, bond and commodity markets working to implement computerized trading procedures might do well to consider the double sealed auction in place of the double continuous model that has served them in face-to-face transactions.

The rapid communication of every bid and offer to every trader requires enormous communications resources. Even in small-scale laboratory markets it is difficult to devise market programs that preserve the flow of bids, offers and contracts that occurs on trading floors. If several hundred geographically dispersed stock traders entered price quotes into a market within a few seconds of one another, the challenge of guaranteeing the proper order of bids, resolving the resulting contracts and disseminating the new market information could be insuperable. In a double sealed auction, batch processing of orders would tremendously reduce demands on the computer network. Precise timing would be less important because the central market computer would effectively synchronize the bids and offers. Price-quantity combinations might be collected from traders once a minute in markets where trading volume was high or as seldom as once an hour where volume was lower.

The Arizona Stock Exchange (AZX), which started trading in March 1992, employs a similar mechanism. Bids and offers are displayed during the interval that the market is open each day, and all trades take place at the equilibrium price when the market closes. About 200,000 shares change hands daily.



DUTCH FLOWER MARKET shows yet another institution for setting prices: the posted price drops with each tick of the clock until a bid is made.

Experimental market research has provided an empirical foundation for tenets of economic theory that were already well established, and it has also yielded insight into the details of how particular rules affect the outcome of the trading process. Thirty years of experiments have also brought good news: under most circumstances, markets are extremely efficient in facilitating the movement of goods from the lowest-cost producers to the consumers who place the highest value on them. Organized exchange thus effectively advances human welfare.

These same techniques are now being applied to help define the shape of markets that do not yet exist. In Southern California, for example, researchers are helping to determine the most efficient rules for trading rights to emit particular pollutants, as required under the amended Clean Air Act. An efficient market in pollutants can reduce control costs because those who can reduce emissions cheaply sell their quotas to those for whom reductions would be costly. Similar work is also being done to develop workable trading rules for markets in natural gas and electricity.

Perhaps the greatest interest in experimental markets comes from the nations of the former socialist bloc, which must repair the economic damage done by decades of totalitarian central planning. This past summer one of us (Smith) helped to organize a seminar in which utility managers and government officials from Poland, Czechoslovakia and Hungary participated in trading exercises under various rules. For people who have never taken part in real markets, the computerized version may be a safe way to hone their skills. Furthermore, as these nations create markets for stocks, financial instruments and commodities ex nihilo, experimental studies could let them develop market institutions that make the most of their scarce economic resources.

FURTHER READING

PAPERS IN EXPERIMENTAL ECONOMICS. Vernon L. Smith. Cambridge University Press, 1991.

COMPUTERIZED LABORATORY EXERCISES FOR MICROECONOMICS EDUCATION: THREE APPLICATIONS MOTIVATED BY THE METHODOLOGY OF EXPERIMENTAL ECONOMICS. Arlington W. Williams and James M. Walker. *Journal of Economic Education* (in press).

EXPERIMENTAL ECONOMICS. Douglas Davis and Charles A. Holt. Princeton University Press (in press).

Meaning and Mind in Monkeys

Nonhuman primates, such as vervet monkeys, seem to communicate in ways that resemble aspects of human speech. But they do not seem to recognize mental states in others

by Robert M. Seyfarth and Dorothy L. Cheney

During the Wimbledon tennis championships in 1981, officials were confronted with an unusual problem. Some male players, notably Jimmy Connors, were regularly grunting loudly as they hit the ball. Their opponents protested, demanding that this practice be stopped. These quieter players claimed the noises were distracting and were emitted deliberately to throw off their timing.

When officials confronted Connors and the other "vocal" players, they received a slightly different explanation. Connors said that some players do grunt on purpose—but not him. He explained that he had no control over his grunting; it just happened when he hit the ball hard. Most of the other grunters were also willing to admit that some players did emit the sounds intentionally, but each denied that he himself had any conscious control over these particular vocalizations.

The Wimbledon officials then observed the different players, trying to discern which grunts were intentional and which were not. They found the distinction virtually impossible to make. The only conclusion they could agree on was that the sounds were indeed distracting, regardless of whether they were made on purpose or just happened as part of the exertion of hitting a ball hard.

The referees' quandary embodies many of the problems confronted by anyone studying the vocalizations of monkeys and apes. East African vervet monkeys, for example, call to one another in a variety of different circumstances. They give loud alarm cries when they see a predator, *wrrs* and *chutters* when they encounter other groups, threat grunts and a different kind of *chutter* when they engage in fights with members of their own group, and quiet grunts during relaxed social interactions. In each case, it is, quite simply, impossible to tell if one monkey deliberately intends to communicate to another or the monkey has no control over its vocalizing and calls out simply as part of ongoing behavior.

In our studies, we have tried to determine whether monkeys have words for things, such as predators, or understand, as we do, that particular sounds represent features of their environment: in other words, do monkeys think? We have also sought to examine whether monkeys have mental states such as knowledge, belief or desire and, perhaps most important, if monkeys do have mental states, whether they recognize that others do as well.

ROBERT M. SEYFARTH and DOROTHY L. CHENEY study social behavior, communication and cognition in monkeys. Seyfarth received a B.A. from Harvard University in 1970 and a Ph.D. from the University of Cambridge in 1977. He has taught at the University of California, Los Angeles, and at the University of Pennsylvania, where he is professor of psychology. Cheney received a B.A. from Wellesley College in 1972 and a Ph.D. from the University of Cambridge in 1977. She has taught at U.C.L.A. and the University of Pennsylvania, where she is professor of biology. Seyfarth and Cheney are currently engaged in a research project on baboons in Botswana.

Scientists who carried out the first field studies of monkeys and apes assumed there were few, if any, similarities between animals' calls and human language. Observers took it for granted that human speech was under voluntary control, that it could be detached from emotion (we can talk about fear without being afraid) and that it involved activity in higher cortical areas of the brain. In contrast, the vocalizations of monkeys and apes were thought to be relatively involuntary, to occur only in highly emotional circumstances and to be under limited higher cortical control. Human words represented objects and events in the external world; the calls of monkeys and apes represented only an individual's emotional state or imminent behavior.

The first hint that these assumptions might be wrong came in the late 1960s from two rather different sources. In 1969 Allen and Beatrix Gardner of the University of Nevada at Reno announced that a chimpanzee named Washoe (pronounced "Wash-show") had learned more than 30 hand signs. Washoe used these signs to communicate about objects, make requests and "just talk." (Researchers later showed that the number of signs an ape can learn is much greater.) To many observers, Washoe was like a circus bear riding a bicycle: training had allowed her to acquire a skill that fell outside the range of a chimpanzee's normal behavior. Others were struck by the large number of signs in Washoe's repertoire and wondered if such communication was completely foreign to chimpanzees or other nonhuman primates.

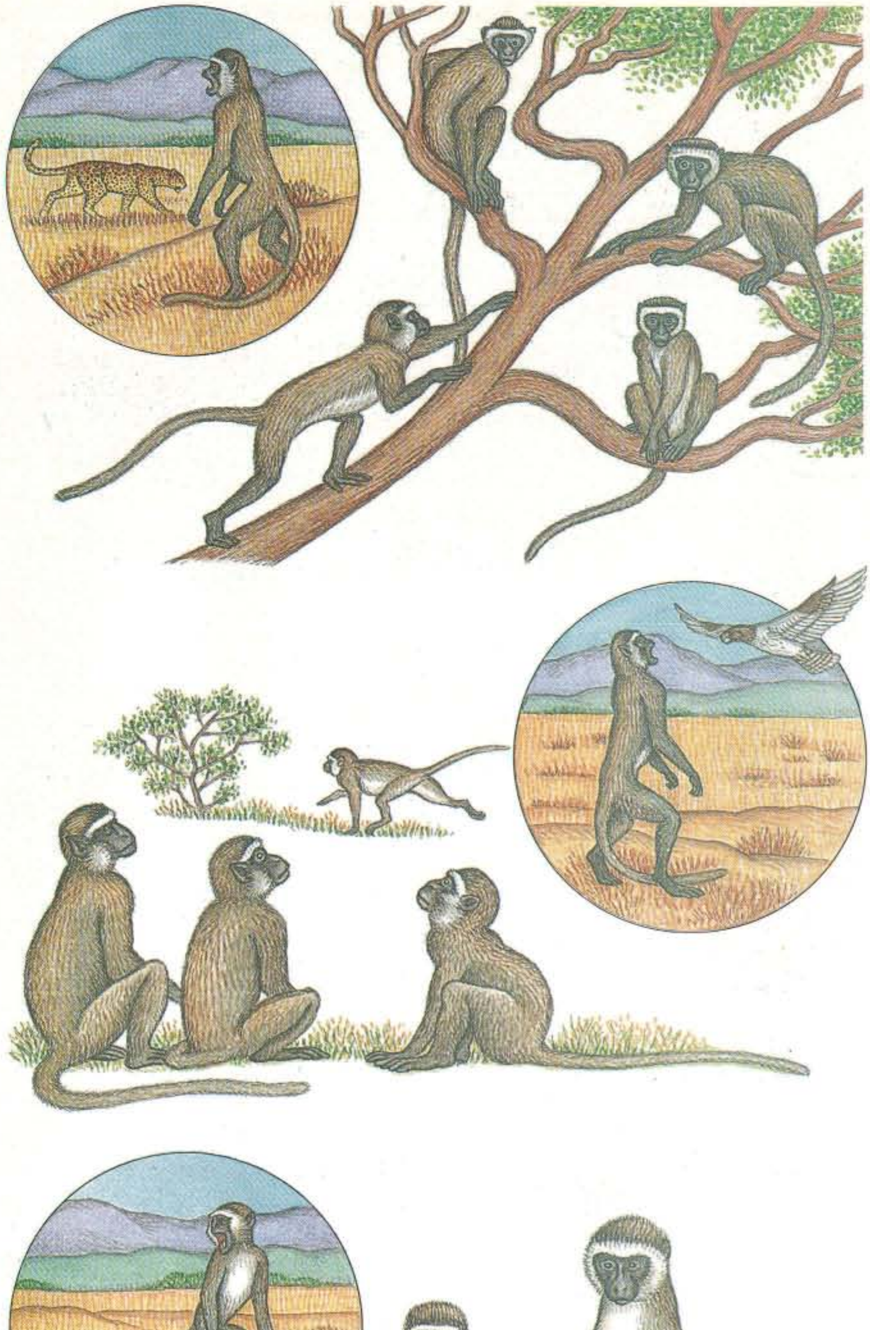
At roughly the same time, in 1967, Thomas Struhsaker, then at the University of California at Berkeley, reported that East African vervet monkeys gave different-sounding calls in response to three predators: leopards, eagles and snakes. Each call elicited a distinct, apparently adaptive, escape response from nearby vervets. Alarm calls given about leopards caused vervets to run into trees, where monkeys seemed safe from feline attack. Eagle alarms caused them to look up in the air or run into the bushes. Snake alarms caused the animals to stand on their hind legs and look into the grass. Like Washoe, vervet monkeys certainly seemed to be using

different signs to denote distinct objects or varied forms of danger.

There were, however, more skeptical interpretations. Instead of indicating different predators, the vervets' alarms might simply be general alerting signals that caused the animals to look all around them. Once monkeys spotted the predator, their responses occurred as a result of what they had seen, not what they had heard. Alternatively, the monkeys' calls might not denote different predators but might instead reflect the relative intensity of fear aroused by leopards, eagles and snakes. If this were true, the analogy between vervet alarm calls and human words would be considerably weakened.

In 1977, as postdoctoral fellows in Peter Marler's laboratory at the Rockefeller University, we designed experiments to test these hypotheses. We conducted our work in the same area where Struhsaker had carried out his

EAST AFRICAN VERVET MONKEYS listen to vocalizations of other members of their troop. The sounds can indicate encounters with another group of vervet monkeys, interactions with members of their own troop or the sighting of a predator.



DIFFERENT ALARM CALLS are given by vervet monkeys in response to the sighting of at least three major predators: leopards (*top*), martial eagles (*middle*) and snakes, such as the African rock python (*bottom*). The monkeys vary their escape route in accordance with the specific call. Experiments conducted with recorded alarm calls show that the cries serve as representational, or semantic, signals for the monkeys.

original study: Amboseli National Park in southern Kenya, at the foot of Mount Kilimanjaro. Here vervets live in groups of 10 to 30 individuals—often including one to eight adult males, two to eight adult females and their offspring. Each group occupies an area of roughly 10 acres that it aggressively defends against incursion by other groups.

As in many other Old World monkey species, such as baboons, rhesus macaques or Japanese macaques, vervet females spend their entire lives in the group where they were born, maintaining close social bonds with their matrilineal kin. The members of each matriline often groom and sleep together as well as support one another in aggressive alliances. Males, in contrast, transfer to neighboring groups when they become sexually mature at about four years of age.

Adult females are ranked in order of dominance, and the offspring acquire a position immediately below that of their mother. As a result, vervet monkey groups consist of a hierarchy of families, with all the members of family A outranking all the members of family B, and so on down the line. Even a male assumes his mother's rank until he transfers to another group. After that move, a male's status depends on more diverse factors, including fighting ability, age and his acceptance by the adult females in his adopted group.

In the early months of our fieldwork we began to accumulate recordings of calls. These vocalizations were given by known individual vervets in encounters with leopards, martial eagles (the monkeys' main avian predator) and pythons. We then played calls to monkeys through a concealed loudspeaker and filmed the animals' responses.

When we examined our results, we found that playback experiments closely duplicated the responses to alarm calls that Struhsaker and we had already observed under natural conditions. Playback of leopard alarms caused animals to run into the trees, recorded eagle alarms caused them to look up into the air or run into bushes and snake alarms caused them to stand on their hind legs and peer into the adjacent grass.

The responses elicited by the recordings of various alarm types argued against the idea that calls were general alerting signals. The sounds seemed to convey information about the presence of specific kinds of danger. Moreover, there was evidence that alarm calls did not simply reflect the caller's degree of fear or excitement. When we altered our tapes to vary the caller's level of excitement by making the calls longer or

shorter and louder or softer, the change had no effect on the qualitatively different responses to each alarm call.

Because we presented alarms when no real predators were around, we could rule out the hypothesis that the monkeys' responses depended on what they had seen rather than on what they had heard. In sum, each type of alarm call, presented on its own, elicited the same response as would a particular predator. We concluded that alarm calls functioned as representational, or semantic, signals.

There are, however, good reasons for caution in drawing parallels between monkey vocalizations and human words. We described the alarms of vervet monkeys as semantic signals because of the way these calls function in the monkeys' daily lives. When one vervet hears another give an eagle alarm call, the listener responds as if it has seen the eagle itself. This behavior suggests that in the monkey's mind the call "stands for" or "conjures up images of" an avian predator even when the monkey has not yet seen the eagle.

But this conclusion could easily be incorrect. For Pavlov's dogs, the sound of a bell may have conjured up images of meat. This fact, however, does not prove that the dogs understood the referential relation between bells and meat in the same way that we understand the relation between, say, the word "chair" and a piece of furniture.

A monkey's call would cease to be a sound and would become a word only if a certain transformation took place. As psychologist David Premack, formerly at the University of Pennsylvania, suggests, this change would occur if the properties ascribed to the call are not those of a sound but those of the object it denotes. Human language offers some excellent examples. If we compare the words "treachery" and "deceit," we typically ignore the fact that the sounds have different acoustic properties. We describe the words as similar because they have similar meanings.

In contrast, the words "treachery" and "lechery," despite their shared sound, are found to be different because they have different meanings. In making these judgments, we recognize the referential relation between words and the objects for which they stand. When comparing words, we judge them to be similar or different on the basis of their meaning, not their acoustic properties. We needed, therefore, to determine if the calls of vervet monkeys qualify as words in this stronger sense in order to claim that monkeys understand the meaning of their vocalizations.

To investigate how vervets compare vocalizations, we borrowed a method from research on speech perception in human infants called the habituation-dishabituation technique. This approach is based on the observation that subjects perceiving the same stimulus over and over again gradually cease responding to it: they habituate. If subjects who have habituated to one stimulus perceive another that they judge to be different, however, the strength of their response increases sharply. The habituation-dishabituation technique thus reveals whether an individual finds two stimuli similar or dissimilar.

To test whether vervet monkeys compare vocalizations on the basis of their acoustic properties or their apparent meaning, we chose as stimuli two calls the monkeys give during territorial encounters with neighboring groups. Monkeys make the first—a long, loud trill called a *wrr*—when they initially spot another group. It seems to alert members of both groups that a neighboring group has been seen. The second vocalization, a harsh, raspy sound called a *chutter,* is emitted when an encounter between groups has escalated into aggressive threats, chases or fighting. *Wrrs* and *chutters* have broadly matching referents—both provide information about another group—but are very different acoustically. Vervets asked to compare *wrrs* and *chutters* in a habituation-dishabituation experiment would therefore judge them to be different if they based their comparison on acoustic properties but similar if they based their comparison on meaning.

To begin our experiments, we selected a subject. On the first day we played a particular adult female's *chutter* to this subject in order to establish the baseline strength of the subject's response. This result was the control condition. On the second day the subject heard the same adult female's *wrr* repeated eight times at roughly 20-minute intervals. Because no other group was present, we predicted the subject would rapidly habituate to the call. Finally, roughly 20 minutes after the last *wrr,* the subject heard the female's *chutter* again.

The logic of our experiment involved playing to our subject the same call, a *chutter,* under two conditions: once in the absence of any prior vocalization (the control) and once after the subject had habituated to the same caller's *wrr* (the test). If our subject judged the *wrrs* and *chutters* to be alike—that is, if monkeys compared calls on the basis of their meaning—then habituation to an individual's *wrr* would also produce habituation to the same individual's *chutter.* In addition, the strength of the response would decrease greatly between the control and test conditions.

But if subjects judged *wrrs* and *chutters* to be different—that is, if monkeys compared calls on the basis of their acoustic properties and not their meaning—then habituation to an individual's *wrr* would not produce habituation to the same individual's *chutter.* In this instance, there would be little difference in the strength of the response between the control and the test conditions.

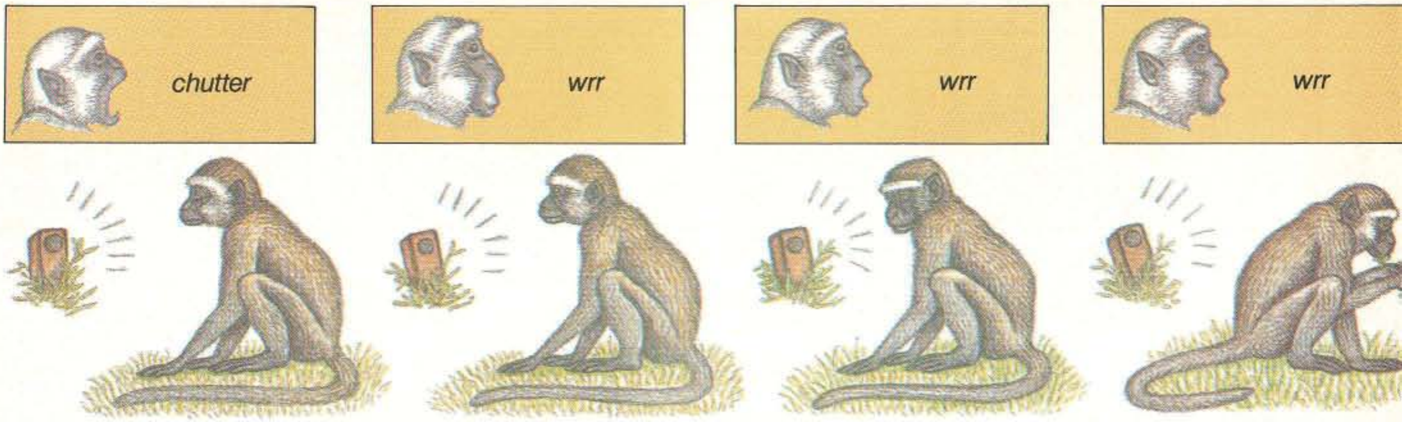
We also decided to evaluate another factor in a monkey's response. From earlier work, we knew that vervet monkeys, like many other birds and mammals, take note of a signaler's identity when responding to calls. Would subjects transfer habituation from one individual to another as well?

In a second series of trials, we varied the test by playing the calls of two group members. On the first day we established baseline data on the strength of a subject's response to individual A's *chutter.* On the second day we played individual B's *wrr* to the subject eight times. After the subject had habituated to B's *wrr,* we tested her to see if she had also habituated to A's *chutter.*

In a third test we examined the effect of changing the call's apparent referent but maintaining the identity of the caller. To do this, we repeated the procedure described for our earlier experiments but instead used as stimuli leopard and eagle alarm calls instead of *wrrs* and *chutters.* Finally, to determine whether subjects would habituate to both individuals and call types, we tested to see if habituation to individual A's eagle alarm would cause subjects to habituate to B's leopard alarm.

Results provide clear evidence that vervet monkeys compare different calls on the basis of their meaning and not just their acoustic properties. When subjects were presented with the same individual's *wrrs* and *chutters,* they transferred habituation to the two types of calls. So if a subject had habituated to individual A's *wrr,* she ceased responding to A's *chutter.* But when subjects heard two calls whose referents were different, they did not transfer habituation to both types of calls: if a subject ceased responding to individual X's leopard alarm call, she responded at normal strength to individual X's eagle alarms.

The identity of the caller also seemed important when vervet monkeys compared vocalizations. Although *wrrs* and *chutters* had similar referents, habituation to individual A's *wrr* did not cause habituation to individual B's *chutter.* Finally, when the calls both had different referents and came from different



HABITUATION STUDIES show that vervets distinguish cries on the basis of their meaning rather than simply on their acoustic properties. *Chutters* and *wrrs* are acoustically different and have slightly different meanings, but both describe encountering another group. When a subject first hears a recorded *chutter* (*far left*), it looks fixedly at the source of the sound. If it later hears the same individual (*yellow*) make a series of *wrrs*, a few of which are shown above (*middle three panels*),

callers, habituation was not transferred.

This series of experiments addressed the question of meaning and reference more directly than did our earlier studies on vervets' alarm calls. The test asked animals to compare two vocalizations, to judge them to be either similar or different and to reveal the criteria the animals used in making their comparison. Our findings suggest that when one vervet monkey hears another vocalize, the listener forms a representation of what that call means. If the listener hears a second vocalization shortly afterward, he or she compares the two calls not just according to their acoustic properties but also according to their meanings.

If we accept the notion that a monkey's call becomes a word when the properties ascribed to the call are not those of a sound but those of what the object denotes, *wrrs* and *chutters* seem to have become words. Thus, vervets seem to have a rudimentary semantic system in which some calls, such as leopard and snake alarms, are quite different in meaning; other calls, such as *wrrs* and *chutters*, are linked to a common referent and can represent shades of meaning within a general class.

Drawing on our knowledge of the vervet's ecology and social behavior, we can suggest why monkeys need such communication and why the cognitive abilities that underlie it may have evolved. If different predators require different escape strategies, then the adaptive value of acoustically different alarm calls is obvious. In addition, monkeys often vocalize when out of sight from one another. This practice favors the evolution of calls whose meanings can be derived from acoustic features alone and that do not depend crucially on contextual cues, such as what the listener sees.

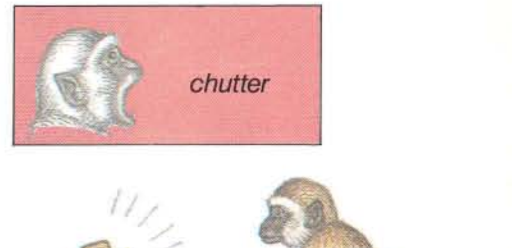
Further, the appropriate response to a vocalization may differ markedly from one individual to the next. A monkey on the ground who gives an eagle alarm may be looking up into the sky, but other monkeys nearby should run into bushes, and those in trees should climb down. These circumstances favor callers who can communicate about events in ways that are relatively independent of their own behavior as well as listeners who can interpret a call's meaning in a manner relatively independent of what the caller is likely to do next. They favor semantic, representational communication.

Our findings, however, did not indicate whether monkey calls could be elevated to the status of human words. Human language involves more than just a referential relation between words and the objects or events they denote. When communicating, we also attribute mental states such as knowledge, beliefs and desires to others. We recognize that mental states have causal power. For instance, we warn young children not to step into the street while waiting for a light to turn green. Our warning is not prompted by their behavior—which may be perfectly safe—but by the lack of knowledge that we attribute to them. As children get older, the same warning provokes outrage and indignation. Older children have learned to attribute mental states to their parents. They know that what parents say reflects what they think, and children do not like being thought of in a condescending way.

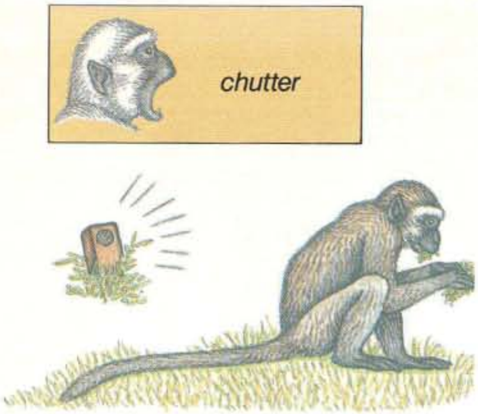
To attribute beliefs, knowledge and emotions to others is to have what psychologists call a theory of mind. We wanted to know whether monkeys could distinguish between their own beliefs and the knowledge and beliefs of others and whether they take special steps to inform an individual who is ignorant or to correct another's false belief.

Hints that animals might be sensitive to the mental states of others came from recent work on what is called the audience effect in various birds and mammals. Marler, now at the University of California at Davis, and his colleagues repeated one of Nikolaas Tinbergen's classic experiments in which a model of a hawk's silhouette was suspended on a wire and "flown" over a group of jungle fowl. Marler found that roosters gave alarm calls in the presence of a hen but remained silent when they were alone with a female of another species. In the wild, Paul Sherman of Cornell University found that adult female ground squirrels were more likely to give alarm calls if they had close kin present than if they did not. In experiments with captive female vervet monkeys, we found that adult females gave more alarm calls when they were with their own offspring than when they were paired with unrelated vervet monkeys of the same age and sex.

Although animals are clearly sensitive



CALLER'S IDENTITY is important to the listener. When one individual (*red*) makes



it habituates and ceases to respond. When the subject finally hears a *chutter* again, it does not respond—the habituation is transferred.

to the presence or lack of an audience, this fact does not prove that they are also sensitive to their audience's state of mind. Moreover, considerable evidence suggests that animals cannot recognize the distinction between an ignorant audience and a knowledgeable one. Roosters and vervet monkeys, for instance, continue to give alarm calls long after their companions have seen a predator and made their escape.

To test for a theory of mind in monkeys, we carried out experiments on two groups of rhesus macaques and two groups of Japanese macaques at the California Primate Research Center in Davis. In their natural habitats, rhesus and Japanese macaques live in groups that are larger but similar in composition to groups of vervet monkeys. By studying monkeys in large outdoor enclosures, we were able to maintain natural social groups while providing some individuals with information that others did not possess.

The subjects for our experiments were adult females and their two- to three-year-old offspring. In the knowledgeable condition, mother and offspring were seated next to each other in a chute that led into a large circular arena. Both watched two scenarios: a researcher placed apple slices in a food bin in the empty arena, or a predator—in this case, a technician who carried a net used to capture monkeys—made threatening gestures and then hid behind one wall of the arena.

In the ignorant condition, mother and offspring were again seated near each other, but a steel partition separated them. Only the mother could see the food bin being filled or the predator hiding. In both knowledgeable and ignorant conditions, after the food had been placed in the bin or the predator had concealed himself, the offspring was released into the arena.

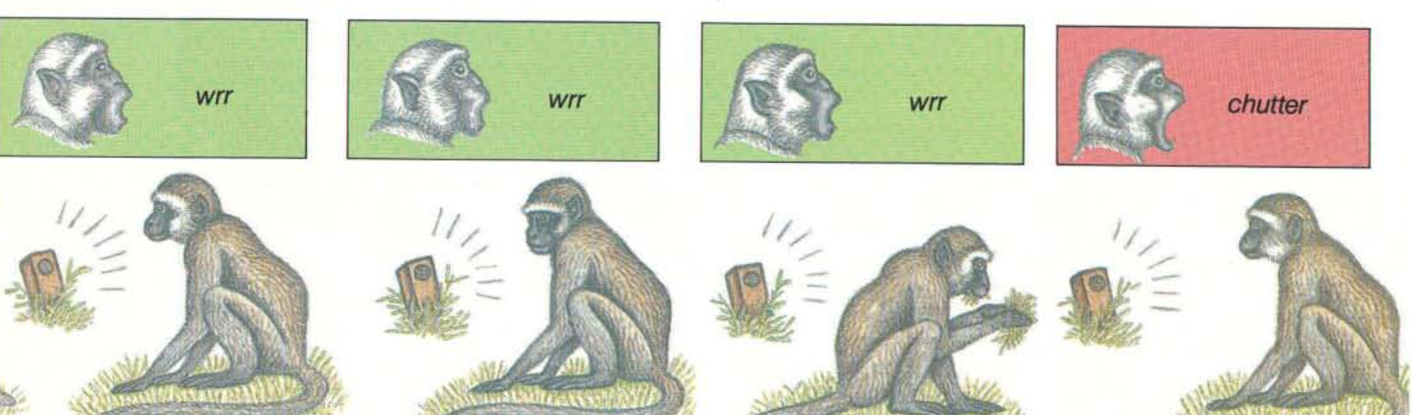
In the ignorant condition the mother had observed her offspring being isolated nearby—and presumably could determine that her offspring was unable to see into the arena. Our goal was to see if the mother, like humans under comparable conditions, would conclude that as a result of its isolation her offspring would not know about the food or the predator. If monkeys are sensitive to others' mental states, mothers should have uttered more calls, or in some other way altered their behavior, when their offspring were ignorant than when they were informed. On the other hand, if animals are unaffected by their audience's mental states, the mother's behavior should have been the same regardless of whether or not their offspring had seen the food or the danger.

As it turned out, mothers had ample opportunity to assess their offspring's knowledge. Ignorant offspring not only were isolated behind a partition, they also behaved differently after being released. In food trials, knowledgeable offspring quickly found the apple slices, whereas ignorant offspring almost never did. In predator trials, knowledgeable offspring stayed close to their mothers after being released, whereas ignorant offspring wandered around the cage, seemingly unaware of the predator's proximity. Despite all these cues, however, we found no difference in the mother's behavior under the two conditions: in neither case did the mother make calls to her young.

Clearly, such negative results do not allow us to distinguish between an inability to attribute a state of mind to others and a failure to make use of such an ability. It is certainly possible that monkeys do recognize the difference between their own knowledge and that of others, but this awareness has no behavioral effect. If rhesus and Japanese macaques can distinguish ignorance and false beliefs in others, however, their apparent failure to act on this information is striking.

Research on chimpanzees suggests that these apes, unlike monkeys, may possess a rudimentary theory of mind. In 1978 Premack and Guy C. Woodruff, then at University of Pennsylvania, conducted experiments in which they showed a chimpanzee named Sarah videotapes of trainers trying to solve a variety of problems. In one case, the trainer could be seen trying to operate a record player whose cord was not plugged into a socket. After each videotape the researchers gave Sarah several photographs, one of which depicted the solution to the problem. She consistently chose the correct photograph.

Premack and Woodruff interpreted Sarah's behavior as evidence that she recognized the videotapes as representing a problem and inferred purpose to the human trainers. Interestingly, when tested with videotapes of a favorite and a less favored trainer, Sarah chose correct solutions for the favorite trainer but incorrect ones for the other trainer. More recently, Daniel J. Povinelli, now at



a *chutter* and a second monkey (*green*) then makes a series of *wrrs*, the subject ceases responding. When the *chutter* of the first individual is replayed, the subject responds again, showing that habituation does not extend to different callers.

YOUNG VERVET MONKEY will learn by observation alone, without tutelage from his mother. In the field and in experiments, mothers have shown an inability to recognize or address ignorance in their offspring. These observations suggest that vervets lack a theory of mind; that is, they do not attribute emotions, beliefs or knowledge to others.

the University of Southwestern Louisiana, and his colleagues obtained comparable results in experiments that required chimpanzees to distinguish between a knowledgeable and an ignorant human condition.

At Gombe Stream Reserve in Tanzania, Jane Goodall once watched an adolescent male chimpanzee, Figan, deceive others in order to obtain a hidden cache of food. As a group of chimpanzees assembled in the provisioning area, Figan suddenly stood up and strode into the woods in a manner that caused all others nearby to follow him. Shortly thereafter, Figan abandoned his companions and circled back to eat the bananas. Of course, anecdotes like this can easily be overinterpreted—one can readily imagine an explanation of Figan's behavior that does not require a theory of mind.

But at the same time, the scientific literature contains an impressive number of cases in which chimpanzees have been observed to deceive others in several different contexts and by an extraordinary variety of gestures, postures and facial expressions. By their number and variety such anecdotes gain in persuasive power and suggest at least the possibility of a theory of mind in apes.

Clearly, the inability to attribute knowledge to others limits a species' actions. Consider some of the initiatives animals cannot undertake if they lack a theory of mind. For example, when infant vervet monkeys begin giving alarm calls or responding to the alarms of others, they make many mistakes. Some, like an infant's eagle alarm given in response to a pigeon flying overhead, are relatively harmless. Other errors, such as that made by an infant who looks up in the air when he or she hears a snake alarm, are more serious, and they actually increase the infant's risk of being taken.

Under these conditions, one might expect adults to intervene and help their infants learn about predators. Somewhat surprisingly, they do not. Despite extensive observations and experiments, we have found no evidence that adults selectively encourage infants who have given alarm calls appropriate to the predators, nor do adults correct infants who have responded to an alarm call inappropriately. Infant vervets learn by observation alone, without explicit tutelage. Such reliance on observational learning is widespread among animals and can, in our view, ultimately be traced to the adults' failure to recognize that their offspring's knowledge is different from their own.

Research on animal communication is inextricably linked to research on how animals think. Our studies of vervet monkeys allow us to conclude that primate vocalizations are not just involuntary shrieks but calls used selectively by individuals who take account of their audience and who use different vocalizations to signal about different features of their environment. Like words, the calls of monkeys function to denote different objects or events, and the monkeys themselves may even recognize the referential relation between a call and the object or concept for which it stands.

At the same time, we are beginning to understand more precisely how the communication and cognition in monkeys differ from that in humans. Although much of human communication is designed to influence the knowledge, beliefs and motives that underlie behavior, there is no evidence at present that monkeys ever communicate with the intent of influencing another animal's mental state. Monkeys and perhaps even apes cannot communicate with the intent to modify the mental states of others because, apparently, they do not recognize that such mental states exist.

FURTHER READING

MACHIAVELLIAN INTELLIGENCE: SOCIAL EXPERTISE AND THE EVOLUTION OF INTELLECT IN MONKEYS, APES, AND HUMANS. Edited by R. Byrne and A. Whiten. Oxford University Press, 1988.

HOW MONKEYS SEE THE WORLD: INSIDE THE MIND OF ANOTHER SPECIES. Dorothy L. Cheney and Robert M. Seyfarth. University of Chicago Press, 1990.

TRUTH AND DECEPTION IN ANIMAL COMMUNICATION. D. L. CHENEY AND R. M. SEYFARTH in *Cognitive Ethology: The Minds of Other Animals.* Edited by C. A. Ristau. Lawrence Erlbaum Associates, 1991.

WHY DO WE AGE?

by Ricki L. Rusting, *staff writer*

To a great extent, the answer is written in our genes. But which ones? New research offers tantalizing clues.

Michael R. Rose, an evolutionary biologist at the University of California at Irvine, sounds like a proud parent when he talks about his extraordinary fruit flies. By successively breeding ever longer-lived females and males, he has created stocks of *Drosophila melanogaster* that he says survive almost twice as long as their standard laboratory-reared counterparts. Longevity is just one of the remarkable traits of these minute Methuselahs. They are, Rose boasts, "superior flies" that are "more robust at every age and better able to resist stress." Even when they are old, many are stronger than ordinary young specimens.

Rose has every reason to be pleased. Profoundly extending life and health in any species is an impressive achievement. Moreover, having developed unusually robust stocks, Rose and his collaborators can now begin to identify the alleles, or variants of genes, that may account for the differences between the superflies and their mundane relatives. One such allele has already been pinpointed.

Rose's fruit-fly work is part of a broader effort to ultimately explain why humans age—why we inevitably deteriorate and die. His findings along with other recent discoveries from a variety of disciplines are encouraging optimism that the question, which once seemed impenetrable, will yield to scrutiny. "We're still groping," says Huber R. Warner of the National Institute on Aging, "but in a dimly lit room instead of in a dark one."

A number of scientists engaged in the struggle say they hope their endeavors will result in ways to slow what they believe is an internally controlled aging process—one that would eventually lead to death even in the absence of accidents, violence and infection. If this intrinsic process does in fact exist and can be slowed, the achievement could help delay or prevent much of the cancer, heart disease and other disabling and deadly conditions to which adults become increasingly vulnerable as they grow older.

Any therapy that retards aging will almost certainly extend life in the bargain, although few scientists seem willing to admit that prolongation of life is a priority for its own sake. "The ideal," says S. Michal Jazwinski of the Louisiana State University Medical Center at New Orleans, "would be to live a long and healthy life and then undergo a rapid demise—to die with your boots on, as they say in cowboy movies."

Rose and others are becoming confident that a full understanding of aging is within reach. "Only in the 1980s have we seen how easy it is to postpone aging," he says. "That's what makes the field very exciting now—we are doing things that work." Indeed, as recently as 15 years ago, most of the available data were descriptive, the notable exception being evidence that caloric restriction could

ELDERLY WOMAN is among the more than a million Americans older than 65 years who live in nursing facilities. Researchers are now beginning to tease out the reasons people generally lose vigor and become prone to an array of disorders in the later years.

prolong the lives of rodents and many other animals. (How diet restriction works is still unclear.) And theories about aging were almost as common as fruit flies. Senescence was caused by a strict genetic program for death. Or by random damage to DNA or to some of the critical enzymes that keep the body running. Or by changes in hormone levels, a decline in the functioning of the immune system or the relentless activity of highly reactive, potentially destructive molecules called oxygen free radicals.

The swarm of opinions prompted Alex Comfort, a pioneer in the field, to write in 1979 that "throughout its history the study of aging ... has been ruinously obscured by theory." Today consensus is still elusive, but investigators generally concur that aging does not have a single cause. Mounting evidence points to a multitude of parallel and often interacting processes, many of them genetically controlled, that combine to ensure eventual decrepitude.

Nature Loses Interest

A number of workers, including Rose, also agree that an understanding of why aging has evolved is a prerequisite to uncovering the physiological bases of senescence. Evolutionary theory, these researchers say, provides a much needed framework for knowing how the body functions and fails—and for identifying the genes involved in controlling life span. "Biochemistry is a valuable part of what we do," Rose says of research on aging in general, "but it doesn't provide an intellectual foundation."

Basic evolutionary theory holds that individuals who are most fit—whose particular mix of alleles best equips them to survive and reproduce in their environment—exert the most influence on the gene pool of future generations. Genetic alterations that enhance fitness (for instance, by enabling one to run faster from danger) will be selected for, or retained, and become common within several generations. Traits are selected because bearers are most likely to survive and reproduce more abundantly, contributing large numbers of offspring to the next generation. Likewise, natural selection will eliminate mutations that are invariably lethal before sexual maturity is attained; thus, affected individuals will leave no offspring to pass along the deadly traits.

Since the late 1800s, many scientists have subscribed to the idea that senescence has arisen because removal of elders increases the fitness of younger individuals, perhaps by reducing the drain on available resources. This scenario suggests that aging is programmed in genes whose sole purpose is to destroy the organism. But the proposition is losing favor, partly because most animals do not survive in the wild long enough to have the chance to become senescent. During the bulk of human history, we, too, died young. The average person born in an industrialized nation today can expect to live about 75 years; for most of our history, life expectancy was closer to 30 or 40 years.

Alternative evolutionary hypotheses hold that genes control senescence but that the unwelcome alleles were not selected specifically for that purpose. The genes of aging—sometimes called gerontogenes—have become ensconced in human chromosomes, these theories assert, because natural selection could not prevent their spread. Alleles that were strictly harmful would persist in a species if the bad effects did not kick in until long after reproduction had begun.

What is worse, as George C. Williams of the State University of New York at Stony Brook proposed in the 1950s, alleles that were destructive late in life would be adopted readily if they somehow improved fitness to any extent early in life—a duality known as antagonistic pleiotropy. Genes that specify the instructions for synthesizing reproductive hormones could well be among those in the traitorous group. Steven Austad of Harvard University posits that the rising risk of breast cancer with age in women might be an example: long-term exposure to the estrogen women require for fertility could predispose breast tissue to malignancy.

Similarly, Caleb E. Finch of the University of Southern California finds that a variety of normal hormones and other regulatory molecules may harm the cells and tissues they influence. The hypo-

NORMAL FIBROBLASTS (*blue*) or connective tissue cells, extracted from human tissue (*left*), become larger as they grow old in a culture dish (*right*). At the same time, their rate of proliferation slows until, ultimately, replication ceases. In-

Selected Changes of Age

Some of the more common physiological and anatomic alterations that occur as people grow older are listed below. Interestingly, certain monkeys and other mammals, including rodents, undergo many of the same changes.

	ORGAN FUNCTION	THERMOREGULATION	IMMUNITY	REPRODUCTION	VISION
CHANGES COMMON BY THE MIDDLE THIRD OF THE LIFE SPAN				Women undergo menopause	
CHANGES COMMON BY THE FINAL THIRD OF THE LIFE SPAN	Ability of heart, lungs and kidneys to function maximally declines	Capacity for coping with changes in environmental temperature becomes impaired	Body's power to combat infection declines; autoimmune responses increase	Male fertility declines	Virtually everyone loses the ability to focus

SOURCE: Caleb E. Finch, University of Southern California

vestigators in several laboratories have now identified many of the genes that seem to control the loss of replicative capacity. What those findings say about how the human body as a whole ages is not yet clear, however.

thalamus and the pituitary gland control ovarian function but also seem to contribute to aging of the ovary, at least in rodents. At the same time, the ovary, which itself sends signals to the hypothalamus and pituitary, seems to promote aging of those organs. Finch further sees these pleiotropic properties as evidence that aging stems to some extent from the activity of—and interactions between—the nervous and endocrine systems.

The genes that enhance early fitness need not be directly destructive late in life in order to cause eventual deterioration. Thomas B. L. Kirkwood of the Medical Research Council in London and Richard G. Cutler of the National Institute on Aging have each proposed that senescence would arise in a population if the body's maintenance systems, which are under genetic control, were good enough to ensure that an organism survived to perpetuate the species but were not able to keep the body going forever.

A Disposable Soma

Kirkwood suggests most animals have not evolved maintenance systems that ensure immortality because doing so would squander energy that could be better put toward reproduction. Since environmental hazards are likely to kill individuals within a reasonably predictable time, a species should invest in protective systems that will guarantee youthful vigor for the expected period but no longer. The rest of an organism's energy supply can then go into maximizing fertility.

In an analogy to industry's practice of investing little in the durability of goods that will be used for only a limited time, Kirkwood calls his model the disposable soma theory. In this case, it is the somatic cells—the nonreproductive cells of the body—that finally are expendable. In contrast, some percentage of germ cells must retain the ability to repair themselves perfectly, or the species would die out.

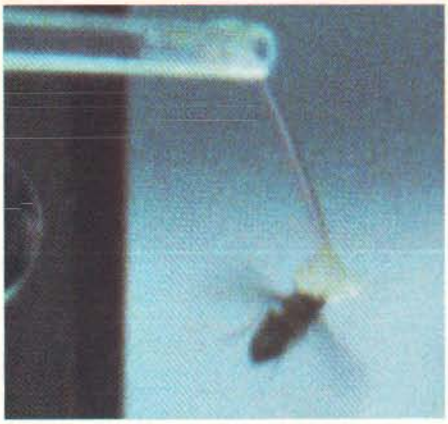
Kirkwood estimates that the investment humans have made in protecting the soma provides enough defensive capacity to assure zest for perhaps 40 years—the period we might have expected to survive during much of our evolution. That investment would include some backup capacity, in case of accidental damage. Presumably, we muddle through mid-life and late life by taking advantage of our reserves. As these backups gradually give out on us, we lose the ability to respond to environmental stress, and, eventually, we die.

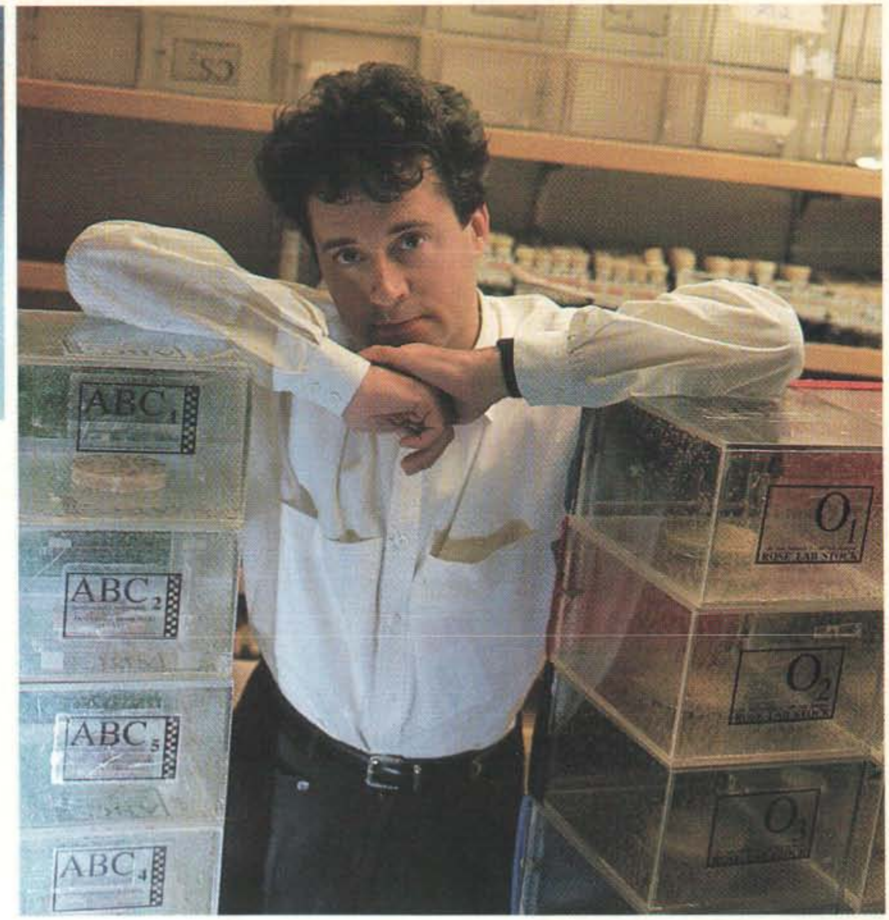
The disposable soma hypothesis lends theoretical support to proposals that aging results from destruction caused by molecules produced in the normal course of living, including oxygen free radicals. The idea that free radicals play a role in aging was put forward in the mid-1950s by Denham Harman of the University of Nebraska. Many biological reactions generate free radicals. Because they carry an unpaired electron, such radicals can oxidize—and thereby damage—DNA, proteins, lipids and other molecules throughout the body. They can also give rise to more radicals and to related oxidants, such as hydrogen peroxide, consequently triggering long chains of destructive activity. Harman proposed that an accumulation of irreversible oxidative damage to cells and tissues throughout the body could result in aging.

The disposable soma concept also accommodates Anthony Cerami's proposal that glucose, the human body's main fuel, is another major factor in aging [see "Glucose and Aging," by Anthony Cerami, Helen Vlassara and Michael Brownlee; SCIENTIFIC AMERICAN, May 1987]. Cerami, now at the Picower Institute for Medical Research in Manhasset, N.Y., has shown that glucose slowly alters long-lived proteins, such as collagen, causing them to become cross-linked, or shackled, to one another. He contends that this glycosylation probably has a part in the stiffening of connective tissue and of the heart muscle that occurs over time.

According to the disposable soma model, these and other wear-and-tear processes could all contribute to aging if they outpaced the capacity of our prevention and repair systems to combat them. But the true test of this and

REACTION TIMES	TUMORS	GROWTH HORMONE	FAT	CORONARY AND CEREBRAL ARTERIES	BONE	JOINTS	LARGE NEURONS IN BRAIN
	Cancers become common in reproductive organs		Storage increases	Some degree of atherosclerosis usually appears	In women, osteoporosis sets in		
Mental and physical responses to specific stimuli become slower		Frequency of secretion declines		Most people have widely disseminated atherosclerosis		Arthritic changes occur	Some grow larger; others atrophy





MICHAEL R. ROSE of the University of California at Irvine has significantly extended the life span of the fruit fly *Drosophila melanogaster* (*far left*) by selective breeding. The insect here, shown in flight, was tethered to fishing line by a drop of Duco Cement for an experiment in which Joseph L. Graves, now at Irvine, demonstrated that the long-lived flies are stronger than normal specimens: they can keep themselves airborne longer under a range of temperatures and humidity levels.

other evolutionary hypotheses will lie in identifying the genes, perhaps hundreds of them, that actually control the molecules that thwart or promote senescence. Many laboratories are now busy attempting to tease out some of the more influential genes.

Genetic Clues

Several investigators—among them Rose, Jazwinski and Thomas E. Johnson of the University of Colorado—are meeting that challenge by seeking genes that can prolong life in relatively simple organisms. "Aging is a puzzle you can understand only if you can compare normal with postponed aging," Rose says. "That tells you what the normal animals are missing. With humans, you don't have a control group."

After Rose created his hardy fruit flies, he and his colleagues compared the proteins made by the experimental and normal insects. One clear difference turns out to be that many of the long-lived flies produce an unusually active version of the antioxidant enzyme superoxide dismutase, which means they harbor a variant of the normal enzyme-specifying gene. Specifically, they produce a highly efficient version of the form of the enzyme typically found in the cytoplasm of cells. In fruit flies, as in humans and other organisms, superoxide dismutases defend against oxidative damage by helping to neutralize a dangerous free radical called superoxide. The genetic difference implies that one reason normal fruit flies age more quickly is that their free radical defenses are not as effective as those of the specially bred flies.

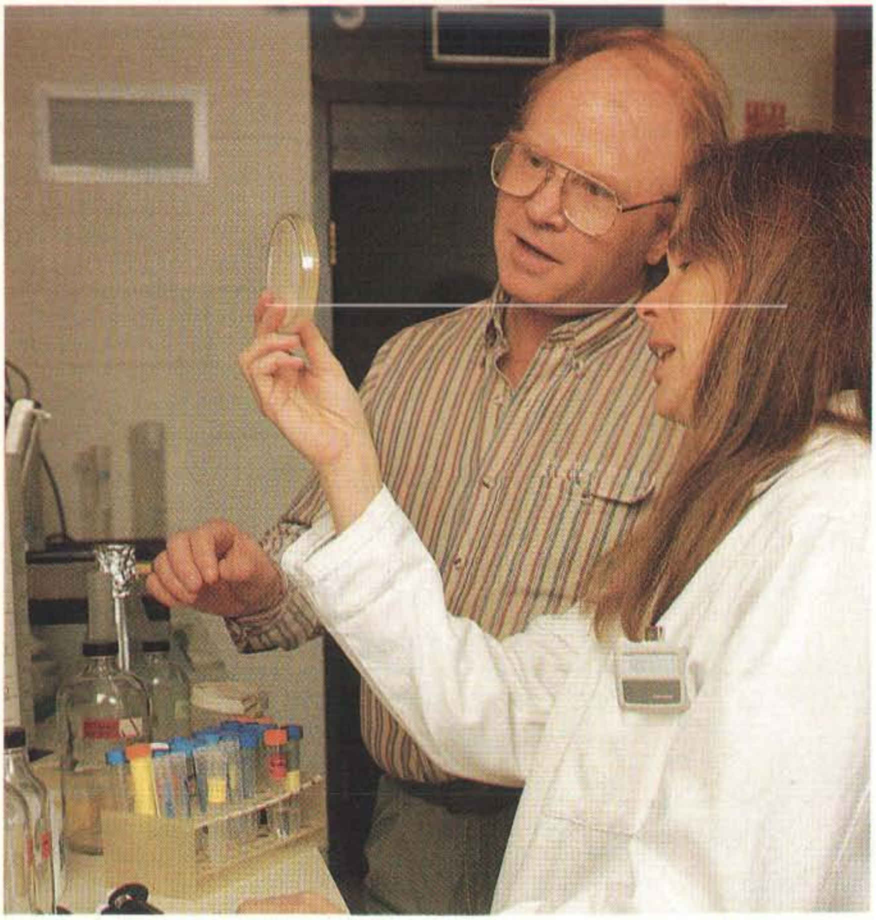
Of course, the variant of superoxide dismutase manufactured by fruit flies is sure to be just one of many factors that influence how quickly such flies age. Rose and Joseph L. Graves and their co-workers at Irvine have found, for instance, that the long-lived flies are more resistant to starvation because they store more fat. (Rose says the insects are so hefty that they spray fat in all directions if they are touched even lightly.) The flies are also less likely to dehydrate, in part because they stow more of the carbohydrate glycogen.

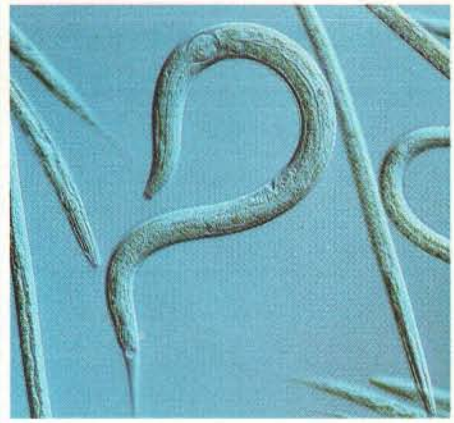
"The work on *Drosophila* is trial-run stuff for doing the same thing in mice," Rose says. "If we can create long-lived mice, specific genes, enzymes and cell processes involved in longevity should be revealed." As mammals, mice are genetically closer to humans than are fruit flies, and so they should have more to reveal about how people age. Mouse research should be more informative, that is, if someone will foot the bill for long-term studies, which Rose estimates would cost about $10 million.

Findings related to Rose's are coming out of Johnson's laboratory at Colorado. Johnson and his colleagues have wielded selective breeding to produce long-surviving versions of a tiny soil-dwelling worm known as *Caenorhabditis elegans*. They have also managed to prolong life in the species by generating random genetic mutations.

Like Rose's group, Johnson's team is attempting to identify the genes that are differentially expressed in long-lived and normal groups (differentially transcribed from DNA into messenger RNA, which is then translated into protein). In 1988 he reported that mutation of a single gene called *age-1* can increase the average life span of *C. elegans* by about 70 percent. Strikingly, the mutant worms produce elevated levels of antioxidants (both cytoplasmic superoxide dismutase and an enzyme called catalase) and are more resistant to the toxic effects of paraquat, a herbicide that leads to generation of the superoxide radical.

The mutation in the *age-1* gene seems to inactivate that gene, which means the encoded protein is no longer made. If the protein's elimination leads to increased production of antioxidants, then it is possible that the normal protein inhibits production of those substances.





THOMAS E. JOHNSON of the University of Colorado, who is examining petri dishes with a technician in his laboratory, has prolonged the life of another multicellular organism: the small, soil-dwelling worm *Caenorhabditis elegans* (*far right*). Johnson and his colleagues achieved the feat in two independent ways—by selectively breeding for longevity and by introducing a mutation in a single gene, called *age-1*. The group is now attempting to clone the *age-1* gene.

Why would an organism deliberately inhibit synthesis of such critical compounds? "I don't think its purpose is to kill the worm at a certain age," Johnson says. Rather he speculates that the inhibition may well be an undesirable effect of some other important function that has not yet been discovered. In other words, he thinks antagonistic pleiotropy is probably operating.

Once Johnson has cloned the *age-1* gene—which he hopes to accomplish soon—he plans to search for a counterpart in mice. If mice harbor a similar stretch of DNA, he might be on the track of a specific gene that could also be involved in human aging. Johnson is fond of a line from *Thus Spoke Zarathustra*, by Friedrich Nietzsche: "You have made your way from worm to man, and much in you is still worm." He is hoping Nietzsche's observation extends literally to the genetics of aging, although he realizes that the causes of senescence in *C. elegans* may be quite different from those in humans.

Jazwinski, aware that we have many genes in common with an even more primitive, single-celled organism, has focused his attention on baker's, or brewer's, yeast (*Saccharomyces cerevisiae*). He has identified several genes that prolong the yeast's life. The best studied of these, *LAG1* (*longevity assurance gene 1*), is more active in young cells than in old ones. Inducing extra *LAG1* activity in older cells, after expression has normally declined, extends their life by about a third. Most important, elderly yeast cells bearing the extra-active gene do not become immortal (as cancerous cells in multicellular organisms do); they simply operate youthfully longer.

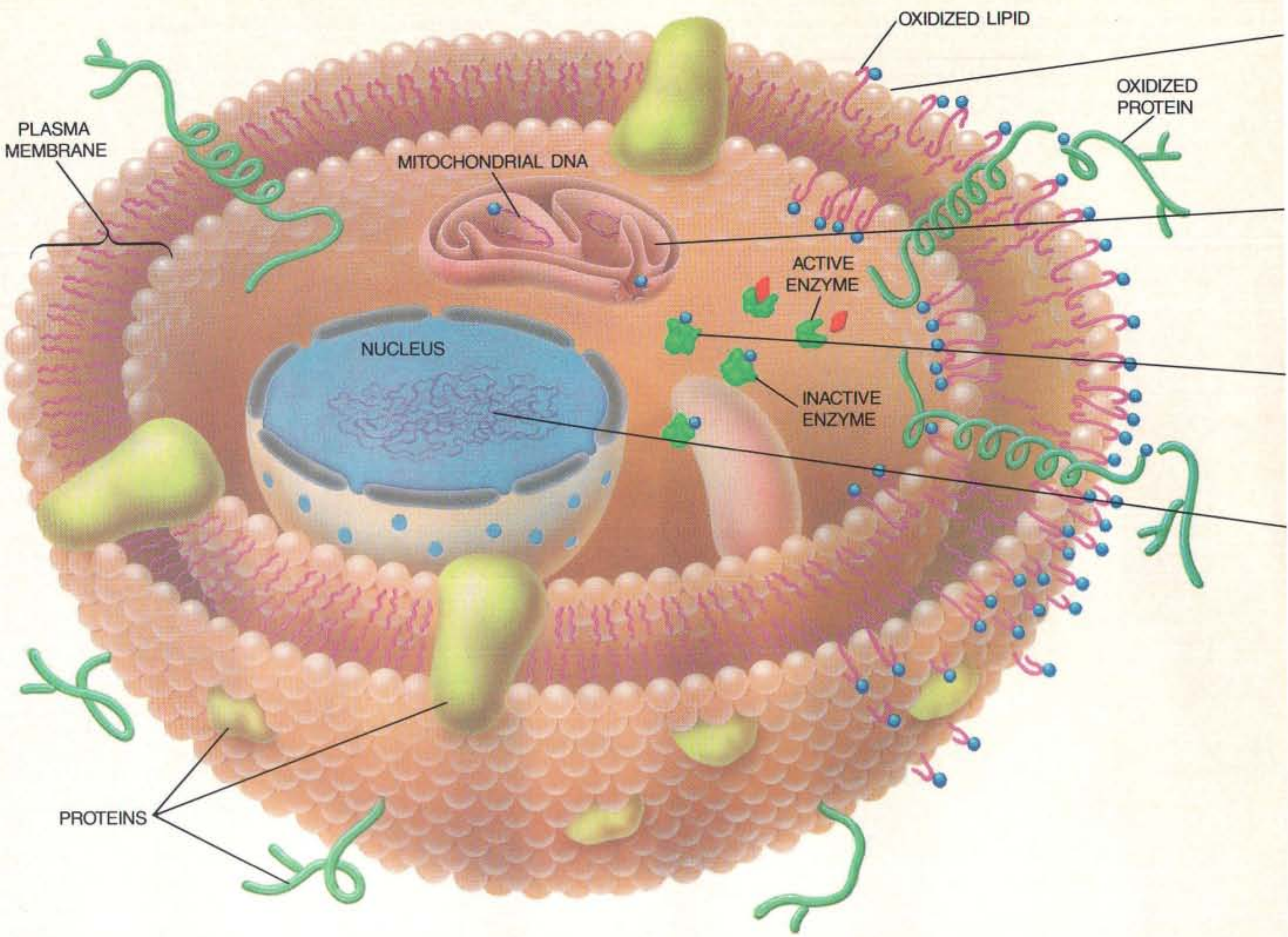
Jazwinski does not know the function of the corresponding protein yet. Nevertheless, he has discovered that a similar gene is expressed in certain human cells. He is now isolating the human gene to determine whether it affects the life span of any human cells. He also intends to overexpress two other longevity assurance genes along with *LAG1* to see whether the combined effects are additive, synergistic or, possibly, destructive. And he expects to evaluate several other genes that his group has reason to believe might influence life span.

Radical Discoveries

Although the function of *LAG1* remains a mystery, the discovery that the antioxidant superoxide dismutase seems to affect longevity in both Rose's fruit flies and Johnson's worms dovetails intriguingly with growing enthusiasm for Harman's free-radical theory. Johnson notes that 25 years ago the theory was accepted as "an interesting hypothesis, perhaps true." Now increasing numbers of investigators are seriously entertaining the possibility that free radicals might indeed be a significant factor in aging.

Much of the evidence for free radicals is still based more on correlation than on clear demonstrations of cause and effect. For instance, if unrepaired damage caused by free radicals were a cause of aging, animals with a high metabolic rate—that is, who burned oxygen relatively quickly—would likely have shorter life spans than would species that consumed oxygen more slowly. The faster metabolizers, after all, would produce free radicals more rapidly. Indeed, the basal, or resting, metabolic rate of a species is inversely related to its life span; mice have a much higher rate than do humans, and they rarely live more than three years.

Cutler of the National Institute on Aging has uncovered other support. He finds that tissues of humans and other long-lived species produce more superoxide dismutase overall and are more resistant to oxidation. Cutler thinks humans age in part because their superior protection against oxidation is nonetheless insufficient to protect them forever. The reactive oxygenated molecules



OXIDATIVE ATTACK on proteins (*shades of green*), lipids (*pink*) and DNA (*lavender*)—indicated schematically by attachment of a small, bright blue sphere—can impair the functioning of many cellular constituents (only some are depicted). One leading theory of aging, originally set out in the mid-1950s, holds that the human body deteriorates over time because it continuously generates potentially destructive oxidizing agents known as free radicals.

are also thought to participate in many disorders common during aging, including atherosclerosis, cancer and arthritis. In several of the diseases, oxidized lipids, which are known to accumulate in aging cells, are implicated.

Some of the most convincing data, however, come from studies of DNA and protein. Until recently, there was little strong proof that DNA was irreversibly damaged over time. For instance, most nuclear genes that have been studied give rise to proteins with the correct amino acid sequences. They do so even though, as Bruce N. Ames of the University of California at Berkeley has estimated, the DNA in each human cell is exposed to some 10,000 oxidative "hits" every day. But studies of mitochondrial DNA indicate that irreparable damage to genes does occur—at least in mitochondria, the organelles that serve as cellular power plants.

Researchers began examining mitochondrial DNA (which specifies about a dozen proteins needed for mitochondrial function) in part because the ability of the organelles to generate energy was found to decline with age. They suspected that free-radical attack on the DNA could be involved because mitochondria are the major source of free radicals in the body and because mitochondrial DNA is particularly vulnerable to oxidative damage. Indeed, the rate of DNA oxidation is much higher in mitochondria than in the nucleus. The genetic material is especially vulnerable because it lacks the histone proteins that bind to and protect nuclear DNA. Moreover, mitochondrial genes are only minimally protected by enzymes that have recently been shown to excise and replace oxidized bits of DNA in the cell nucleus, according to Kelvin J. A. Davies of Albany Medical College.

These observations raised the possibility that a relentless oxidative assault on mitochondrial DNA could slowly interfere with the functioning of mitochondria in normal aging adults. When enough of the organelles became seriously harmed, the cells they inhabit would become starved for energy. And when enough of the cells in an organ were impaired, the function of the organ would decline.

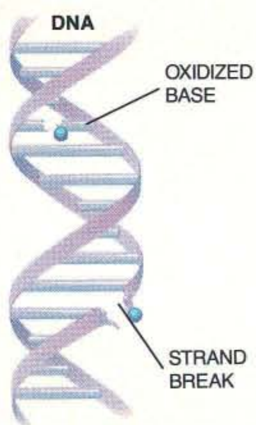
Consistent with this scenario, Gino A. Cortopassi and Norman Arnheim of the University of Southern California reported in 1990 that mitochondrial DNA in heart and brain cells of aged adults carries a defect not found in fetal tissue. And, according to Douglas C. Wallace of the Emory University School of Medicine, it appears likely that a significant number of mitochondrial DNA molecules may be defective in elderly people. Wallace also speculates that several chronic diseases common in old age might be related to mitochondrial failure, including late-onset diabetes and Parkinson's and Alzheimer's diseases. "The trouble," Cutler says, is that "you can still ask, what is the proof that

MEMBRANES
Oxidation of lipids and proteins can undermine the integrity and tightly regulated permeability of both the plasma membrane and the membranes surrounding internal organelles. Damaged cells may fall apart.

MITOCHONDRIA
When these organelles, which are a major source of free radicals, are themselves severely damaged, cells can starve for energy. Mitochondrial DNA and membranes are particularly prone to oxidation.

ENZYMES
Many enzymes can be inactivated by oxidation at their active sites. Attack by free radicals can also damage other intracellular and extracellular proteins.

CHROMOSOMES
Free-radical attack on the DNA molecules that constitute chromosomes can potentially interfere with the accuracy or amount of protein made by a cell.



these changes cause any problems? We don't have it." But researchers working on proteins may have some of the proof Cutler is seeking.

During the late 1980s, investigators developed a technique for measuring carbonyl groups, which are a common by-product of protein oxidation. Subsequent measurements revealed that the amount of oxidized protein in various human cell types increases exponentially with the age of the donor. Striking evidence was also obtained from examining fibroblasts—cells that produce collagen and other constituents of connective tissue—taken from patients with rare genetic disorders that cause premature aging. Those suffering from progeria and Werner's syndrome showed levels of carbonyl groups dramatically higher than in normal subjects of equal age.

On the basis of such findings Earl R. Stadtman of the National Heart, Lung and Blood Institute estimates that as

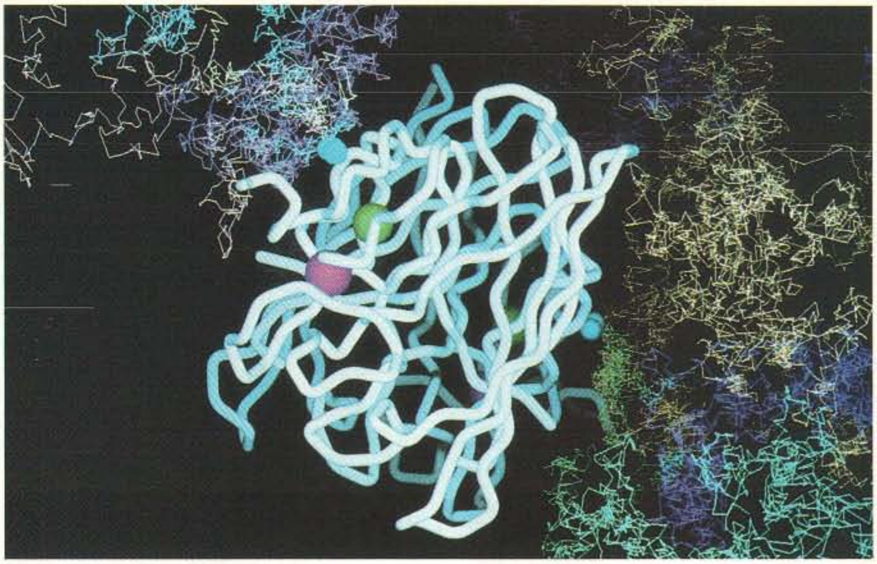
Defenses against Oxidative Damage

A variety of defenses prevent or repair molecular damage caused by free radicals, but as a group they are thought to be imperfect. Some evidence indicates that certain of the defenses also become less effective over time.

	CLASS	MOLECULE	ACTIVITY
ANTIOXIDANTS (neutralize free radicals or otherwise limit their activity)	**ENZYMES**	Superoxide dismutases	Convert the superoxide radical $(O_2^{\cdot -})^*$ to hydrogen peroxide (H_2O_2)
		Glutathione peroxidases and catalases	Convert hydrogen peroxide to harmless water and molecular oxygen (O_2)
	OTHER	Vitamin E and beta carotene	React with free radicals, thereby preventing them from attacking cellular constituents; are fat soluble and so can protect membranes
		Uric acid and vitamin C	React with free radicals in the cytoplasm
		Metal chelators	Prevent iron, copper and other transition metals from catalyzing oxidation reactions
REPAIR SYSTEMS (degrade, repair or replace damaged molecules)	**PROTEIN REPAIR**	Proteinases	Cleave oxidized proteins
		Proteases	Cut the products of proteinase activity
		Peptidases	Chop up products of protease activity; amino acids can then be recycled to make new proteins
	LIPID REPAIR	Phospholipases	Cut out damaged parts of oxidized lipids in membranes so that other enzymes can repair injured segments
		Acetyltransferases	Thought to replace fatty acids cleaved from lipids
		Glutathione peroxidase and transferase	Help to repair oxidized fatty acids without ejecting large segments from membranes
	DNA REPAIR	Exonucleases and endonucleases	Clip out damaged segments of DNA
		Glycosylases and polymerases	Fill in the gaps left by exonucleases and endonucleases
		Ligase	Seal the repairs

*The dot next to O_2 in the superoxide radical indicates the presence of an unpaired electron.

SOURCE: Kelvin J. A. Davies, Albany Medical College



TYPE OF SUPEROXIDE DISMUTASE (*whitish structure at center*) active in the cytoplasm of human cells has been altered slightly by scientists to enhance its activity. The mutant more rapidly guides dangerous superoxide radicals (*blue spheres*) toward copper (*green*) and zinc (*purple*) atoms that participate in superoxide breakup. The colorful squiggles flanking the enzyme trace the trajectories followed by separate superoxide molecules en route to their demise. Superoxide dismutase is now used to reduce inflammation and is under study for other applications, potentially including amelioration of degenerative diseases that become common late in life. This version of the enzyme is one of several variants that have been created.

much as half of the protein, including many enzymes, in elderly individuals might be oxidatively damaged—and thus nonfunctional. That degree of oxidation, he adds, would almost certainly promote senescence. "I find it unacceptable," he insists, "to believe we can lose 50 percent of the activity of enzymes without it having a rather deleterious effect on metabolism."

Last year John M. Carney of the University of Kentucky Medical Center and his co-workers offered the first direct evidence that free-radical attack on proteins impairs physiological function. First, they showed that the level of oxidized protein in the brain increases as gerbils age and that treatment with a compound called phenylbutylnitrone (PBN) can reduce oxidation to youthful levels. Then, in an as yet unreplicated experiment, they administered PBN to elderly gerbils for two weeks. Before treatment, older animals tested in a maze exhibited impaired short-term memory compared with younger ones. At the end of the treatment period, the elders performed at the level of youngsters (although without continued treatment, performance later deteriorated).

Stadtman, Davies and others have evidence indicating that oxidized, inactive proteins accumulate in cells because the body's ability to degrade them declines. Davies says other cellular repair systems seem to become impaired as well over time or less able to respond strongly to oxidative stress. What exactly impedes the repair machinery is unknown.

The free-radical data thus support the possibility that we age because of an inability to cope perfectly with wear and tear, not because our chromosomes include a program designed to kill us. But the idea of a specific death program has not yet been laid to rest. Indeed, one prominent line of inquiry is quite consistent with it.

Aging in a Dish

That line of research centers on investigations of cultured human cells. In the 1950s dogma held that human cells capable of proliferating in the body would replicate indefinitely in cell culture. If that were true, it meant people aged and died not because of some intrinsic program for cellular decay but because of processes originating outside of cells, at some higher level of physiological organization.

The dogma toppled when Leonard Hayflick and Paul S. Moorhead, then at the Wistar Institute, reported in 1961 that normal human fibroblasts have a built-in limit to the number of times they can proliferate. More specifically, there is a set number of times—roughly 50—that populations of fibroblasts taken from a new embryo can double.

The existence of what is now called the Hayflick limit has since been confirmed many times. Hence, analyses of why human cells in culture cease to proliferate might provide clues to the failure of the organism as a whole. "By understanding why cells stop proliferating, we may be able to understand something about aging," says James R. Smith of the Baylor College of Medicine. But, he adds, "I firmly believe it is going to be a lot more difficult—probably at least 1,000 times more difficult—to understand aging in the intact human being."

The likelihood that aging of cells in vitro is related to aging of the body has gained backing from several discoveries. It turns out that doubling capacity declines progressively with the age of cell donors. Similarly, fibroblasts from patients afflicted with Werner's syndrome cannot replicate as many times as cells from normal subjects of equal age. Also, all human cell types studied so far have characteristic limits to proliferation in culture.

In the past several years, investigators have identified genes whose expression systematically changes as fibroblasts in culture lose their doubling capacity—just as would be expected if there were a genetic program for aging. Workers have evidence that genes on chromosomes 1 and 4 participate in this loss of replicative ability. And Judith Campisi, now at the Lawrence Berkeley Laboratory, has determined that shutting off of the gene c-*fos* precedes and probably influences many of the other changes that have been identified. When c-*fos* is silenced for good, fibroblasts fail to duplicate their DNA. Consequently, they also fail to divide.

Smith suspects that c-*fos* may itself be finally quieted by a still more primary process. "What I think is going on," he speculates, "is that cells produce an inhibitor that blocks initiation of DNA synthesis." He says he thinks he has now cloned the gene for the inhibitor, so he should soon be able to put that idea to a test. He quickly acknowledges, however, that his conception is one of many possible explanations. "We're still not to the point where we can separate cause from effect, because the inhibitor's activity might be a secondary effect of something else."

One candidate for that "something else" is shortening of telomeres, the long stretches of DNA that cap both ends of every chromosome and prevent them from decaying. Calvin B. Harley of McMaster University in Ontario, Carol W. Greider of Cold Spring Harbor Laboratory and their colleagues have found that the length of the telomeres progres-

sively decreases in somatic cells that replicate in the body and decreases as well during the aging of fibroblasts in culture. Moreover, telomere length is a better predictor of division potential than is the age of cell donors.

Presumably, the shortening occurs because the machinery responsible for duplicating DNA during cell division has a strange flaw. It eliminates a small bit of each telomere in every new copy of DNA it makes [see "The Human Telomere," by Robert K. Moyzis; SCIENTIFIC AMERICAN, August 1991]. These findings imply that the telomere could be the clock that determines loss of proliferative capacity in cells. Interestingly, Harley and Greider have established that telomere length is maintained, or even slightly increased, in sperm and in transformed, or immortalized, cells. That maintenance could help explain why normal germ cells and malignant cells do not lose replicative capacity.

One could make the case, then, that the body decays when organs inevitably stop being able to replace damaged cells. On the other hand, critics charge that people do not die because fibroblasts stop doubling; the cells usually have plenty of divisions left in them when their "owner" perishes. Skeptics also point out that studies of replicative senescence cannot shed much light on the processes leading to deterioration of nondividing cells, notably neurons and heart muscle cells. Those cells function admirably for years.

Rose is among the critics. He calls the work on cell culture "technically beautiful" but insists that the investigators are, in fact, "spinning their wheels." They are probing cell proliferation and its limits usefully, he says, but they have yet to show that their results reveal anything about how people age. "If these cell biologists really were discovering anything important about organismal aging, they would be able to use their findings to postpone aging," he adds. "But in 30 years of work, this they have never done."

Smith thinks such criticisms are unjust. The genetic changes seen in fibroblasts may represent only one aspect of the aging process, he says, but that aspect is nevertheless potentially significant. "What might be going on is that there are local areas in which cells are not functioning properly and cannot be replaced," he notes. He cites the single-celled lining of endothelium in blood vessels: "If the endothelial cells in a small area of the blood vessel had lost the ability to proliferate and were either lost or dysfunctional, that could lead to the processes that eventually cause atherosclerosis." Impairment of proliferative capacity, he adds, appears to be a problem in the immune system.

"I'm not sure we die of one precise thing," Smith contends, "but I think the loss of cell proliferation may well be an important contributing factor." Vincent J. Cristofalo of the Medical College of Pennsylvania, a major advocate of the in vitro studies, agrees. In a presentation at a recent conference on the molecular biology of aging, he pointed out that cells growing old in culture exhibit similarities to cells undergoing senescence in the body. In addition to losing replicative capacity, he asserted, they are larger and show changes in the structure of the nucleus. "I would be hard-pressed," he said, "to argue that the cells do anything but age by any definition of aging."

How can researchers reconcile the fibroblast data with evolutionary hypotheses? Many workers, including Smith, maintain that a limitation on proliferative capacity may have evolved not as a death program but as a defense against cancer. The loss of proliferative ability, then, might be yet another example of antagonistic pleiotropy. The "program" that helps us to resist cancer also prevents immortality. Although many people eventually do fall ill with cancer, our built-in resistance renders us significantly less prone to tumors than many other animals.

In Search of a Unified Theory

In spite of the tremendous progress by researchers studying cultured cells, free radicals, longevity-determining genes and other promising avenues, the aging process in humans is still largely a black box. Adding to the confusion is the fact that age-related changes do not occur uniformly, either among individuals (who differ in their susceptibility to specific disorders and in exposure to environmental stresses) or among cells. "The larger message," Finch says, "is how remarkably selective particular age changes are at the molecular, cellular and organ level in any species."

Nevertheless, Finch believes a clear vision of the aging process will emerge. "My hunch," he says, "is that over the next 20 to 30 years, the study of aging mechanisms will create a crowning intellectual synthesis of currently separate biomedical approaches to human disease."

Among the most hopeful are investors eager to back the development of new drugs and treatments that may be tantamount to the fountain of youth or, for the moment, partial fountains. For instance, new companies are investigating the possibility that PBN and different types of superoxide dismutase might serve as remedies for various disorders common late in life.

"There is a scramble out there right now," Warner says. Venture capitalists call him constantly, he adds, "to get the latest ideas on aging and make millions."

Exactly how far life might be extended by tinkering with the aging process is anyone's guess. Jazwinski, for one, is on record as speculating that the maximum life span (the oldest age anyone can achieve), which is currently about 120 years, might go as high as 400 years. But even he thinks the average Joe who receives treatment, say, 30 to 50 years from now, is more likely to gain about 30 extra years of life.

Of course, prolonging survival even that much could have profound social, economic and environmental consequences. This possibility raises an issue that seems to be receiving little systematic attention right now—namely, should life be extended?

Not surprisingly, few leading researchers express much doubt that it should. Johnson says he plans to organize serious discussion of the subject, although his instinct is that people might treat the earth more respectfully if they knew they would be around to suffer the effects of their polluting ways. And Finch, echoing the words of others, says, "We're already doing it" by watching our diets and trying to keep active.

Rose, typically, offers one of the more provocative answers. "If I compare prolonging life with building better bombs or putting a man on the moon or any of many things scientists can do with money, this is far superior—though it is not better than vaccinating all the children of the Third World."

FURTHER READING

VITALITY AND AGING: IMPLICATIONS OF THE RECTANGULAR CURVE. James F. Fries and Lawrence M. Crapo. W. H. Freeman and Company, 1981.

MODERN BIOLOGICAL THEORIES OF AGING. H. R. Warner, R. N. Butler, R. L. Sprott and E. L. Schneider. Raven Press, 1987.

LONGEVITY, SENESCENCE, AND THE GENOME. Caleb E. Finch. University of Chicago Press, 1990.

MOLECULAR BIOLOGY OF AGING: PROCEEDINGS OF A UCLA COLLOQUIUM, HELD AT SANTA FE, N.M., MARCH 4–10, 1989. Edited by Caleb E. Finch and Thomas E. Johnson. Wiley-Liss, 1990.

REVIEW OF BIOLOGICAL RESEARCH IN AGING, Vol. 4. Edited by Morton Rothstein. Wiley-Liss, 1990.

EVOLUTIONARY BIOLOGY OF AGING. Michael R. Rose. Oxford University Press, 1991.

Christmas in the House of Chaos

John Tractor brought home the perfect fractal pine for his family's Christmas tree. Each branch was an exact, reduced copy of the whole pine. A beautiful and fragrant fractal, indeed. Inside the tiny Tractor residence, chaos reigned, as it always does.

John's wife, Jane, had covered herself in bread crumbs as she tried to prepare her grandmother's recipe for recursive turkey: turkey stuffed with sausages stuffed with miniature turkeys stuffed with even tinier sausages stuffed with— His daughter was racing around the kitchen with a large butterfly net, screaming, "Stop the hurricane!"

John, who could tolerate chaos but not pandemonium, scolded his daughter. "Lorenza Tractor, why don't you settle down, like your brother, Pointer?"

"But, Dad, I'm trying to stop the butterfly-effect hurricane," Lorenza said, panting.

"What?"

"You know. A butterfly flaps its wings in Japan, and a month later there's a hurricane in Brazil."

John pursed his lips. "But this isn't Japan. Or Brazil," he added.

Lorenza gave him a withering look. "Dad, you can't tell where the hurricane will hit or which butterfly is going to do the damage, so I may as well hunt one here."

"In December?"

"You've got a point, Dad," she conceded. "Hey, may Pointer and I help decorate the Christmas tree? Pointer can climb to the top of the tree, and I'll throw the glass ornaments up to him."

"Absolutely not," Jane cried from across the room.

"Aw, Mom—"

"Let's make some decorations," Pointer said.

"That's a good idea," Jane said in a relieved tone.

"Yeah, we can get a whole pot of glue and some paint and—"

"Why not start by designing some decorations?" Jane suggested quickly. "You've got that new graphics package for the computer and a color printer."

"Great!" Pointer yelled. "Let's do some chaos decorations for Christmas."

"That's a neat idea," Jane agreed. "But I really want something with a bit of pattern to it. I need some objects to hang from the ceiling, kind of like lanterns and stars, but you ought to be able to think of something more original. And some unusual wrapping paper—but I want it to have a pattern as well as being chaotic."

"Crumbs, Mom. You can't make order out of chaos."

Jane sighed. "It's what I spend all my time doing. Off you go."

Grumbling under their breath, Lorenza and Pointer traipsed up the stairs to the attic, where the computer sat. "Order and chaos in the same design," Lorenza muttered. "Crazy. Mom's just trying to get us out of the way so we can't cause trouble."

"We always do," Pointer admitted.

"True. Pointer, I'm not going to be defeated. If Mom wants order and chaos at the same time, that's what she'll get. Chaotic patterns—Can you think of things that have patterns?"

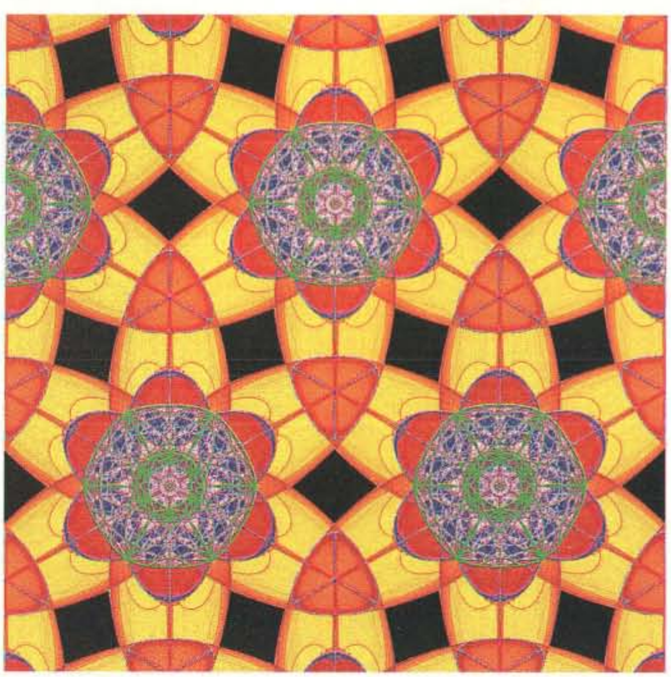
"Wallpaper," Pointer said. "Usually it has rows and rows of roses or things; but we could stamp a fractal design in each square of some grid. That would combine order and chaos."

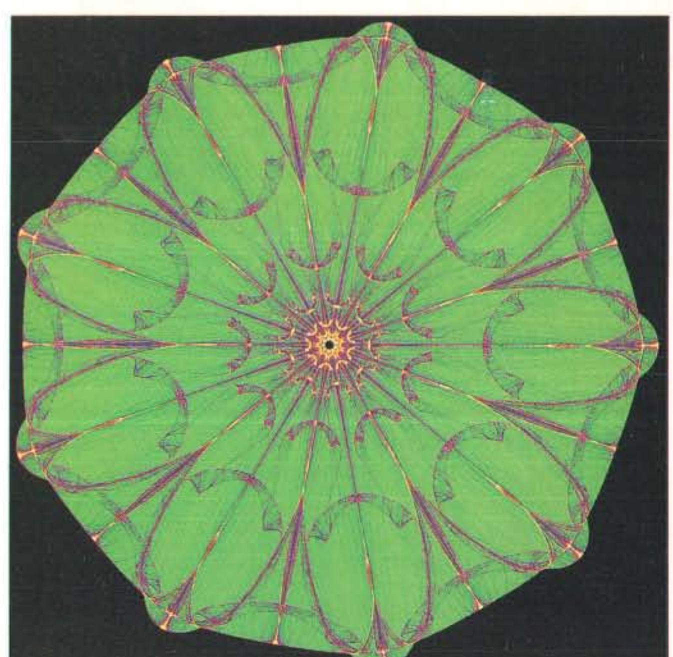
"Mmmm—it's cheating a bit though."

"Okay, then, kaleidoscopes. They make patterns out of anything. Remember when we put all those bits of colored rock under a kaleidoscope?"

"That's a brilliant idea! We'll put bits of chaos under a kaleidoscope instead. Where are those fractal postcards? I'll scan them into the computer while you

SYMMETRY AND CHAOS can be combined to produce exquisite patterns.





program the mirrors."

The children set to with enthusiasm, but after half an hour, Pointer pushed the postcards aside. "No, it doesn't look right. It's like my wallpaper idea—it's really just a lot of copies of the same bit of chaos."

"Yes, and you can see the boundary where the kaleidoscope mirrors are joined together. We need a process that doesn't produce boundaries in the first place. Of course, I know. Symmetric chaos."

"I don't think I've come across that."

"It was invented only recently by Michael Field in Australia and Martin Golubitsky in Texas. How does it go? Oh, right. Let's start with ordinary chaos. The way that works is that you iterate a mapping of the plane and follow where the points go."

"You mean you choose some fixed rule, some formula that turns pairs of numbers into pairs of numbers. You choose a starting pair, then apply the rule, then apply the rule to the resulting pair, and so on."

"Precisely, Pointer. If you think of the pairs as coordinates of a point in the plane, you can plot the successive points on the computer screen. The first few hundred points are 'transients'—that is, they are produced before the system settles down to its long-term behavior. So we won't plot those. The remaining points build up a shape. It is known as an attractor because no matter where you start, the points are attracted to that region, and you end up with the same picture."

"So in some cases, the attractor should be very dull—a single point, for instance. But in others, we should get wonderful swirling forms with lots of intricate detail. Chaos!"

"Right," Lorenza said. "Symmetric chaos is exactly the same, but there's one extra trick: you make the mapping symmetric."

"I don't understand what you mean by that. I can see when an object is symmetric—but a mapping?"

"Well, we can begin with the idea of symmetry. A symmetry of an object is a way to move it so that afterward it looks exactly the same. You can rotate a square through angles of 0, 90, 180 and 270 degrees, and if people turned their backs while you did it, they wouldn't notice any change."

"What about 360 degrees?"

"That has the same effect as no rotation at all. Every point of the square ends up where it started. But squares have four more symmetries, reflections in mirror lines that run horizontally, vertically and down the two diagonals. That makes eight symmetries altogether."

"Gotcha."

"Now, suppose you have a mapping from the plane to itself—a rule that sends any particular point to some other specified point. For instance, the rule might be that every x coordinate is cubed and every y coordinate is cubed. Such a rule would be written as $(x,y) \rightarrow (x^3,y^3)$. The rule creates an image of the original."

"Okay."

"When I say that such a mapping has a symmetry, I mean that symmetrically related points have symmetrically related images. Suppose the symmetry is 'rotate by 90 degrees,' and I take any point P and apply the mapping to get a point P'. Then I rotate P by 90 degrees to get Q and apply the mapping to get Q'. I want Q' to be P' rotated through 90 degrees. If it always is, then I say that the mapping has 90-degree rotational symmetry."

"That's quite a mouthful! Can I see an example?"

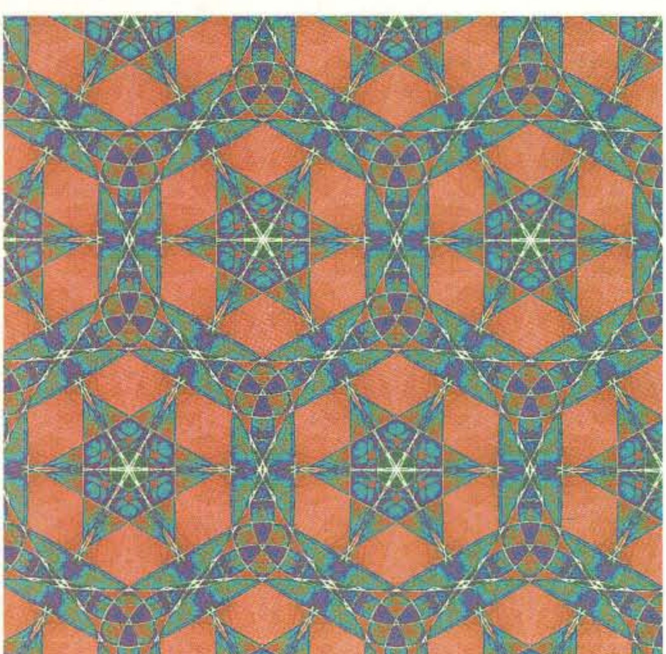
Lorenza pursed her lips in thought. "Yes, let's take a look at the 'cube' mapping, namely, $(x,y) \rightarrow (x^3,y^3)$, and see whether it has square symmetry. Suppose P is the point (x,y). Then P' is (x^3,y^3). The 'rotate through 90 degrees' symmetry is also a mapping, and it sends (x,y) to $(y,-x)$, which is Q. Then the 'cube' mapping sends Q to Q', which is $(y^3,(-x)^3)$.

"The question is, is Q' the result of rotating P' by 90 degrees? Well, if we rotate P' through 90 degrees, we get $(y^3,-x^3)$. That is the same as Q', because $(-x)^3 = -x^3$."

"Okay, I think I get it now. If you had used a mapping like $(x,y) \rightarrow (x^2,y^2)$, you wouldn't get rotational symmetry, because $(-x)^2$ is not the same as $-x^2$."

"That's it. You have to check all the appropriate symmetries, of course— not just one of them. If the mapping is to have a mirror symmetry, then points that are mirror images of one another have to map to points that are mirror images of one another in the same mirror. For instance, reflection in the y axis sends (x,y) to $(-x,y)$. You can check that the 'cube' mapping has that symmetry, too."

"I suppose," Pointer commented, "that for threefold or fivefold symmetry, like equilateral triangles and pentag-



Formulas for Symmetric Chaos

Readers can write their own computer programs to generate images of symmetric chaos. The program starts with some arbitrary point (x,y), which represents some location on the computer. Then, to create an image with three-fold symmetry, for example, the program should implement the following "mapping":

$$x \to ax + bx(x^2+y^2) + c(x^3-3xy^2) + d(x^2-y^2)$$
$$y \to ay + by(x^2+y^2) + c(3x^2y-y^3) - 2dxy$$

These instructions simply mean that the program should enter the initial values for x and y into the equations on the right side of the arrows to obtain a second point; the program should then substitute the new values into the formulas to obtain a third point and so on. The programmer is free to choose the values of a, b, c and d, but beware that many choices lead to pictures that are isolated dots or fall off the computer screen. Try $a = 1.89$, $b = -1.10$, $c = 0.17$, $d = -0.79$, for example.

After 100 or so iterations of the mapping, the program should begin to keep track of how many times the points hit a given area of the screen. Each area is then assigned a color depending on how many times it was hit. Gradually, a chaotic image with threefold symmetry should appear on the screen. Experiment and be patient!

Mapping with fourfold symmetry:

$$x \to ax + bx(x^2+y^2) + c(x^4-6x^2y^2+y^4) + d(x^3-3xy^2)$$
$$y \to ay + by(x^2+y^2) + c(4x^3y-4xy^3) + d(-3x^2y+y^3)$$

Mapping with fivefold symmetry:

$$x \to ax + bx(x^2+y^2) + c(x^5-10x^3y^2+5xy^4) + d(x^4-6x^2y^2+y^4)$$
$$y \to ay + by(x^2+y^2) + c(5x^4y-10x^2y^3+y^5) + d(-4x^3y+4xy^3)$$

Quilts:

$$x \to a\sin(2\pi x) + b\sin(2\pi x)\cos(2\pi y) + c\sin(4\pi x) + d\sin(6\pi x)\cos(4\pi y) + kx$$
$$y \to a\sin(2\pi y) + b\sin(2\pi y)\cos(2\pi x) + c\sin(4\pi y) + d\sin(6\pi y)\cos(4\pi x) + ky$$

Here x and y must lie between 0 and 1; the programmer is free to choose a, b, c, d and k, but k must be a whole number. Nearly all choices lead to nice pictures, but the program should plot several copies of the picture side by side to produce the full effect.

onal stars, the same goes for rotations through 120 degrees or 72 degrees."

"Precisely," Lorenza said. "Although the formulas for the mappings get more complicated when the rotational angles aren't right angles. Now, suppose you iterate a symmetric mapping by choosing a starting point and applying the mapping over and over again. And suppose the resulting attractor is chaotic, which it often is. Then you get lots of nice disorder. But you must also see some effects of the symmetry. In fact, a lot of the time the attractor is symmetric. If the mapping has fivefold symmetry, say, then you get attractors that look like five-pointed stars, but with lots of symmetrically placed chaotic detail."

"You said 'a lot of the time' the attractor is symmetric. Why not always?"

"Well, sometimes you get several distinct but symmetrically related attractors. Five mirror-symmetric attractors in a pentagonal formation, say. If there's only one attractor—which is often the case—it has to be symmetric.

"Field and Golubitsky did one more thing. They counted how many times the point being iterated hits a particular place in the attractor and color-coded the picture accordingly. Say you use red for places that get hit between one and 100 times, blue for between 101 and 200 times, and so on. The coloring gives you an idea of how probable it is that the iterated point will be in a particular region. That gives attractive symmetric and chaotic colored designs. They're like chaos in a kaleidoscope, but because the symmetry was built into the mapping, you don't see artificial-looking edges. A chaoleidoscope, so to speak."

"Let's try it. What's the formula for the mapping?"

"I'll have to look it up [*see box at left*]. It's a bit complicated."

Several hours passed while Lorenza and Pointer produced multicolored chaotic patterns. Unlike the designs that are made by a kaleidoscope, these were not formed by making lots of copies of the same image, rotated through various angles or reflected in various lines. Instead the shapes formed gradually, emerging out of the dark background like trees out of the mist. New points seemed to be added at random, yet somehow the mapping seemed to "know" that its attractor had to be symmetric.

Lorenza and Pointer were so quiet as they worked that eventually Jane and John peeped in to see what mischief they were up to. The children proudly displayed their designs.

"Oh, they're wonderful," Jane declared. "But can you make me some wrapping paper, too?"

"Of course," Lorenza exclaimed. "Field and Golubitsky call them 'quilts.' They're chaotic attractors with the same kinds of symmetry as wallpaper—a lattice symmetry, where you can slide the picture sideways and fit it on top of itself. They can also have rotational and reflectional symmetries, if you want. They're done in just the same way, but you have to use mappings with the same kinds of symmetry that wallpaper has. And you have to draw several copies of the pictures side by side, like tiles on a wall, to see the full effect."

"I'm amazed," John said. "I wouldn't have thought you could mix order and chaos in the same object—and I certainly didn't think the result could be so beautiful. Why don't we decorate the tree now?"

The children jumped up in excitement, grabbed a box full of paper decorations and ran downstairs. By the time John and Jane caught up with them, half of the Christmas tree had been decorated, and the room was strewn with paper. John whispered to Jane, "Actually, I guess I always knew order and chaos could be beautiful."

FURTHER READING

FEARFUL SYMMETRY: IS GOD A GEOMETER? Ian Stewart and Martin Golubitsky. Blackwell Publishers, 1992.

SYMMETRY IN CHAOS. Michael Field and Martin Golubitsky. Oxford University Press, 1992.

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION (required by 39 U.S.C. 3685). *1A.* Title of publication: Scientific American. *1B.* Publication number: 509-530. *2.* Date of filing: September 22, 1992. *3.* Frequency of issue: monthly. *3A.* Number of issues published annually: 12. *3B.* Annual subscription price: U.S. and its possessions, 1 year $36, all other countries, 1 year $47. *4.* Complete mailing address of known office of publication: 415 Madison Avenue, New York, NY 10017. *5.* Complete mailing address of the headquarters of general business offices of the publisher: 415 Madison Avenue, New York, NY 10017. *6.* Full names and complete mailing address of publisher, editor and managing editor: Publisher, John J. Moeling, Jr., 415 Madison Avenue, New York, NY 10017. Editor, Jonathan Piel, 415 Madison Avenue, New York, NY 10017. Managing Editor, Michelle Press, 415 Madison Avenue, New York, NY 10017. *7.* Owner: Scientific American, Inc., 415 Madison Avenue, New York, NY 10017; Holtzbrinck Publishing Holdings Limited Partnership, c/o Scientific American, Inc., 415 Madison Avenue, New York, INC 10017: (a) Holtzbrinck Publishing Group, Inc. (General Partner), 100 West 10th Street, Wilmington, DE; (b) Georg von Holtzbrinck GmbH & Co. (Limited Partner), Gaensheidestrasse 26, D-7000 Stuttgart 1, Germany. *8.* Known bondholders, mortgagees and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages or other securities: none. *9.* For completion by nonprofit organizations authorized to mail at special rates (DMM Section 424.12 only). The purpose, function and nonprofit status of this organization and the exempt status for federal income tax purposes: not applicable to this organization. *10.* Extent and nature of circulation: *A.* Total number of copies (net press run): average number of copies each issue during preceding 12 months, 845,587; actual number of copies of single issue published nearest to filing date, 971,688. *B.* Paid and/or requested circulation: *1.* Sales through dealers and carriers, street vendors and counter sales: average number of copies each issue during preceding 12 months, 138,580, actual number of copies of single issue published nearest to filing date, 228,538. *2.* Mail subscription (paid and/or requested): average number of copies each issue during preceding 12 months, 507,390; actual number of copies of single issue published nearest to filing date, 507,454. *C.* Total paid and/or requested circulation (sum of *10B1* and *10B2*): average number of copies each issue during preceding 12 months, 645,970; actual number of copies of single issue published nearest to filing date, 735,992. *D.* Free distribution by mail, carrier or other means of samples, complimentary and other free copies: average number of copies each issue during preceding 12 months, 19,612; actual number of copies of single issue published nearest to filing date, 28,000. *E.* Total distribution (sum of *C* and *D*): average number of copies each issue during preceding 12 months, 665,582; actual number of copies of single issue published nearest to filing date, 763,992. *F.* Copies not distributed: *1.* Office use, left over, unaccounted, spoiled after printing: average number of copies each issue during preceding 12 months, 13,243; actual number of copies of single issue published nearest to filing date, 47,460. *2.* Return from news agents: average number of copies each issue during preceding 12 months, 166,762; actual number of copies of single issue published nearest to filing date, 160,236. *G.* Total (sum of *E*, *F1* and *2*—should equal net press run shown in *A*): average number of copies each issue during preceding 12 months, 845,587; actual number of copies of single issue published nearest to filing date, 971,688. *11.* I certify that the statements made by me above are correct and complete. (Signed) John J. Moeling, Jr., Publisher.

BOOK REVIEWS by Philip and Phylis Morrison

Science Books for Young People

This year brought a sharp change of emphasis among the books: we found readers who enjoy activity—out-of-doors or in the lab—particularly well served.

Technology

HUMANPOWER: CARS, PLANES, AND BOATS WITH MUSCLES FOR MOTORS, by Roger Yepsen. Macmillan Publishing Company, 1992 ($14.95).

Aimed at kids in the middle grades, this book opens with a quick history of human-powered vehicles, centered on the bicycle. Roger Yepsen reminds us that there were two times of fast progress in this domain: the Golden Age of the safety two-wheelers in the 1890s and again today. In between, human engines deferred to the new, hot, noisy, little box, the gasoline engine, a "barn full of power" for a sip of fuel.

It is fine to see the 1930s Velocar of François Faure. The first recumbent racing bicycle, really laid back, was so fast that it was quickly barred from official races. They are here again, designs both recumbent and fully streamlined, doing all that can be done to reduce air drag. These new human-powered vehicles have raced for years at 50 miles an hour and might become workable around town yet still cheap enough to find a public. For now, recumbency is within reach—expect to pay 50 percent more than for an ordinary bike of similar quality—but well-made streamliners cost at least 10 times more.

Muscle-powered boats are the oldest of all, as canoes, kayaks and rowboats evidence. An eight-oared shell is handsome and fast, but no faster than the watercycle pedaled by six Frenchwomen a century and more back.

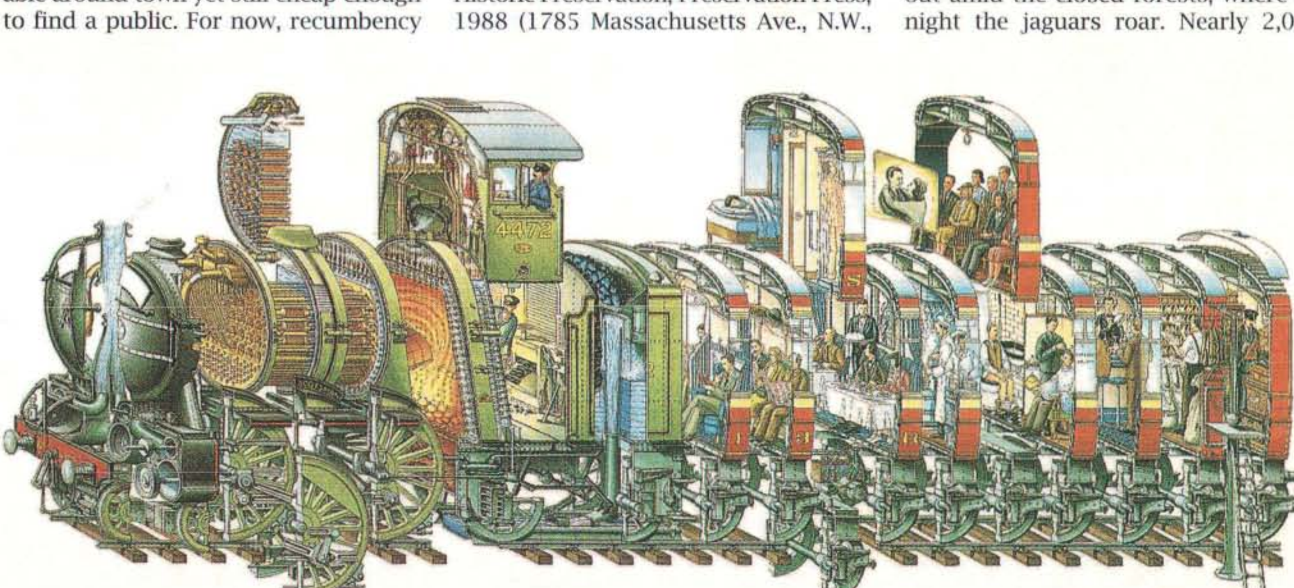
Flight is the last challenge to humanpower, first met in 1977 by Paul MacCready's huge, fragile gossamer aircraft. "Practical" flight by pedal seems beyond us: not enough power per pound, except on the moon. The best our author offers young readers is to study the basics of science, so as to prepare to join the college-student engineers now working so well on pedal power. This is a readable and informed summary, with a fine list of books to take you faster and farther.

STEPHEN BIESTY'S INCREDIBLE CROSS-SECTIONS. Illustrations by Stephen Biesty, text by Richard Platt. Alfred A. Knopf, 1992 ($20). **SAN RAFAEL: A CENTRAL AMERICAN CITY THROUGH THE AGES,** by Xavier Hernández. Illustrated by Jordi Ballonga and Josep Escofet. Translated by Kathleen Leverich. Houghton Mifflin Company, 1992 ($16.95). **WHAT IT FEELS LIKE TO BE A BUILDING,** by Forrest Wilson. National Trust for Historic Preservation, Preservation Press, 1988 (1785 Massachusetts Ave., N.W., Washington, D.C. 20036) (paperbound, $10.95).

These three unusual books differ sharply in topic and intended readership. All the same, they are linked; they all look toward architecture, if not always straight at that art itself, and all present their subjects through vivid images.

Biesty takes us right inside 18 wonders of our world, ancient and modern, by slicing them like loaves of bread. Annotation crowds his detailed and colorful cross sections. Start with his castle: cutaways show you archers at their high slits, the bailiff at the accounts, guards underground at the door of some poor wretch in his oubliette. End with the space shuttle orbiter and its payload, six or eight high-tech slices packed with plumbing and tankage, worth long study. Chartres Cathedral, the 1930s steam train the *Flying Scotsman,* and the *Queen Mary* herself suggest the range of these wonders, made fully credible by a British artist of incredible energy and skill.

San Rafael is an "invented metropolis," a convincing fiction displayed to us in 14 detailed bird's-eye views in black-and-white. It shows cross sections not in space but in time. The site lies on the banks of a plausible tropical river, flowing slowly from the foothills of the Guatemalan highlands down to the plains of Yucatán. We see it first in 1000 B.C.; only a few cornfields stand out amid the closed forests, where by night the jaguars roar. Nearly 2,000



years later we see it as a splendid Mayan city, pyramids, temples, ball courts, rich marketplace. In a few centuries, that city vanishes back into the conquering jungle; the jaguar roars again around the overgrown stones. Then from fortress to colonial cathedral square, to bullring and sugar mills and a railway once owned by the General Fruit Company, we watch San Rafael rise again, named in a new tongue. There it is today, old splendor become restored archaeology, busy streets with banks and a university, the ball court magnified into a soccer stadium bigger than the cathedral, quiet farm villages around the city replaced by chaotic slums. With each of the aerial views are small, detailed sketches of the people and how they lived. This social history in image puts a hard question to serious young readers: Will their century temper the harsh divisions and heedless changes in the San Rafaels around the world?

Forrest Wilson's big pages hold at most a few lines of text, so they invite young readers. The strong and simple pictures spread in black-and-white their eye-trapping solid forms and sharp line. Columns do bend; we see them bending, not only ordinary columns but tall, thin human shadow figures. A heavyset acrobat at the bottom holds up seven men, each one standing on the shoulders of the one below. The higher the man, the lighter his build. An arch is all squeeze; once more, human figures endure the stresses you too can feel. Buttresses push in like butting goats. Squash, squeeze, droop, tug, bend or brace: any building is clearly a matter of strong verbs. The human images are convincing in this memorable demonstration that like our buildings we too feel and challenge unsleeping gravity, not just in metaphor but in brick and in brawn.

STRING: TYING IT UP, TYING IT DOWN, by Jan Adkins. Charles Scribner's Sons, 1992 ($13.95).

Jan Adkins learned the ropes as a sailor, and he does not forget his marlinspike seamanship, fussy because at sea the lines must work. But he is mindful of what small-stuff string can do as well. He offers a guide to hauling sheets of plywood on the car roof, to shoelaces, buttons and holding your hat in any wind, and on how to make boats secure at dockside, decorate packages prettily, tie a bow tie and safely swing a comfortable hammock. Of course, he explains the terms of the art and a little of how line is made and stored and safely thrown to a partner, all with a light heart.

HOW THINGS WORK, by H. Richard Crane. American Association of Physics Teachers, 1992 (5112 Berwyn Rd., College Park, MD 20740-4100) (paperbound, $18).

Everyday steel ball bearings, not very costly, are precise "in diameter and sphericity" to better than one part in 10,000. How can such elegant spheres be made so cheaply? Only one shape, the true sphere, every diameter the same, can maintain its form indefinitely under prolonged random abrasion. The procedure is a patient use of abrasive powder on rough balls of steel held between two circularly grooved steel disks set into relative rotation. It is essential that the balls do not roll or slide repetitively but steadily reorient themselves and switch among grooves. Randomness is assured by a clever scheme for steady, slow circulation of the balls outside the disks and back in again. At last the balls are all worked to one single diameter. The manufacturers comment that the chance precision of the process exceeds the precision with which they can measure the product.

All that was a matter of knowing consultation with the right experts. But Dick Crane is not the usual compiler of ingenuities. He is a distinguished physicist at the University of Michigan at Ann Arbor, a versatile, consummate experimenter of tireless curiosity, deep insight and wide experience. Not many of these 70 one- or two-page accounts of mysteries old and new are so passive as his satisfying account of ball bearings.

How do elevator call "buttons" work—if they don't move at all when you push? What about supermarket bar code readers or automatic focus—six kinds—in your camera? How do hydraulic-lift canal locks manage to lift huge loads over and over without any motors or pumps? How do fluorescent lights work?

Crane is an adept of "reverse engineering," dissection and analysis of any puzzling gadget. Somehow get inside; reduce detail to principles and tough physics and clever design to key concepts; never ignore quantity where it really counts. He levels with the reader as few experts do. Even his best reverse engineering could not explain what happened in the electronic stud finder with its glowing diodes. He found its pulses on his radio and gathered that it was the dielectric effect of the wooden stud itself that counted. Just how that was used escaped him down an integrated chip. (It was a subtle matter of changing pulse lengths.)

These wonderful pieces are columns collected from a decade of a monthly journal, *The Physics Teacher,* that is aimed at high school teachers and open to technically curious high school readers. The book is a treasure for any school with a flow of physics students, who will learn more from this true investigator than the answers.

Animals

IN THE BEARS' FOREST, by Bertil Pettersson. Translated by Steven T. Murray. Color photographs by the author. R&S Books, 1991 (Distributed by Farrar, Straus and Giroux) ($11.95).

Here is the bears' forest, all right. No sign of human beings is to be seen in any of these pictures, taken in a Swedish forest, except for the little soundproof bear shed of boards scented with pine resin to conceal the human smell. There the author spends a month every spring waiting for bears to come, his camera ready. He built the shed at forest edge, at the end of a path into a bog, and he hid there in springtime for years before he saw any bear close by. The big, wary bears live most of the year deep in the huge forest, very hard to find. They sleep all winter. But in spring when the earliest berries ripen in the valley, and in the fall when berries grow ripe on the nearby mountain slopes, here you can see the bears without a long and dangerous stalk through the forest.

Spring nights in far northern Sweden glow with the late-setting sun until midnight, so the camera can catch bears, even though they never come before eight o'clock. Bears are not everyday visitors; "sometimes a few weeks may pass without bears." Powerful males have moved quickly and smoothly right up to the shed. "If the bear is less than twenty yards away, I slide the peephole

window shut." Wonderfully a mother and her three cubs came again and again. Pettersson was able to watch the frisky, playful cubs grow to handsome young bears on their own in the forest, and he shows them to us beautifully year after year.

Such quiet, patient intimacy with bears is a surprise to Americans. The grizzlies in Yellowstone Park once became beggars for food from crowds of humans; those animals are slowly growing wild again. Meanwhile, the author reminds us, grown-up grizzlies are much faster, much stronger, and have far more endurance than people. "If you meet a bear, talk to him and walk away calmly. It's no use running or climbing up a tree."

ELEPHANTS CALLING, by Katharine Payne. Crown Publishers, 1992 ($14).

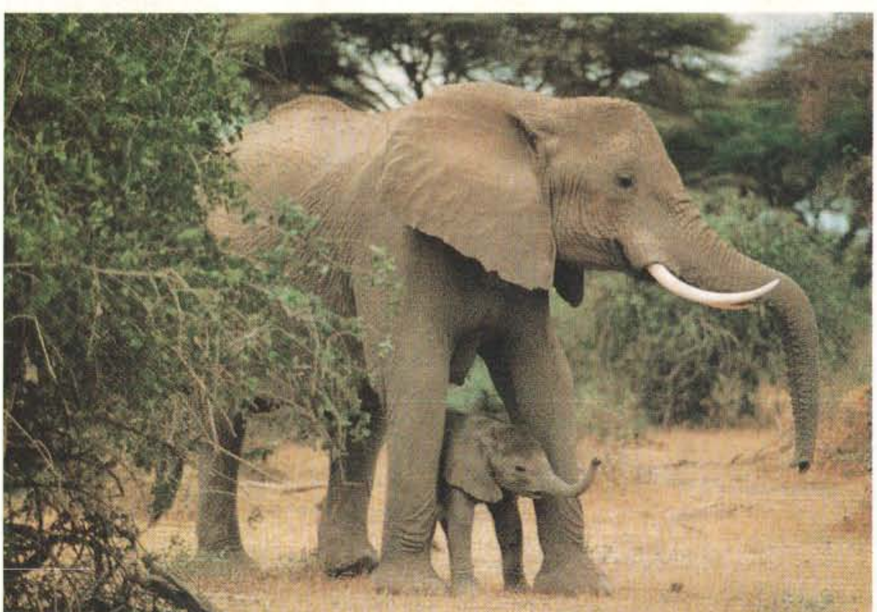
This winning book is a pleasure for anyone old enough to enjoy reading or being read to. A color photograph or two and a few direct sentences on every page give life and meaning to all we see. Simply and warmly we meet two fascinating individuals in context: little Raoul, a baby African elephant, and a talented human being, the author herself.

Payne is a "musical biologist," widely known for her discovery that wild elephants hear and call in infrasound, using tones so low in pitch that humans cannot usually hear them. As a child she sang in the church choir right in front of the organ pipes that sounded the deepest notes. Sometimes she would feel the air "throb and flutter" when the note sounded was at the edge of what human ears can hear.

About 10 years ago she felt the same kind of throb while she was near the elephants' cages in a zoo. Could it be that elephants make powerful sounds lower than the deepest organ pipe? Just so. Now she and her partners study wild elephants and their signals, carrying recording equipment that lets people hear the deep calls they would miss without electronic aid. Patiently and quietly, they follow one elephant family, eager to find out what life is like for animals with "a voice that travels over the horizon" to other elephants miles away. That is how they happened to find wobbly Raoul with his coat of reddish fuzz almost as soon as he was born one sweet January morning.

Many elephants on the Amboseli reserve below snowy Mount Kilimanjaro now bear human names like Raoul because the biologists have identified and named them one by one. Raoul and his mother, Renata, belong to one family group, all given R names. There are 14 of the R family now, including the wise old grandmother, aunts and great-aunts, "taking care of each other and their babies," and six young cousins. Elephant fathers do not regularly live with any family, and the young brothers all leave for good after about 14 years, but the females remain lifelong.

The biologists followed the elephants in a small truck for several years, never getting out of it while within sight of the animals. The whole herd came to regard the familiar truck as part of the landscape; elephants would scratch against it as if it were a rock. About a month after Raoul's birth, the time came for the thirsty R family to make its way across the dusty plain toward the swamps where the water never fails. All across the plain, other elephant families were doing the same thing. Suddenly the family stopped in their tracks. "Then Renata bellows," and all ran forward to greet their old friends, the F family, at the swamp edge. They had heard and exchanged infrasound calls the unhearing biologists had missed.



To understand animals, you must "watch for a long time without disturbing them. Then gradually your mind opens to what it would be like to have different eyes, different ears... different needs, different fears, and different knowledge from ours." This is a lesson in the science of behavior even deeper than infrasound, and the basis of Katharine Payne's plea that we recognize how much ivory costs—the entire future of these great-voiced and peaceful societies.

Math

FROM ZERO TO INFINITY: WHAT MAKES NUMBERS INTERESTING, by Constance Reid. Fourth edition. Mathematical Association of America, 1992 (paperbound, $19).

Strictly fulfilling her title, Reid offers a dozen enticing chapters, the first 10 of them labeled by the digits in increasing order. The eleventh is called after the decisive noninteger that Leonhard Euler named *e*. The closing chapter is named for one special infinity (the smallest of all Georg Cantor's transfinite numbers), dubbed aleph-zero.

The simplest arithmetical instances and the biggest number-fishes caught by computer are here alike in the 1992 version of a graceful book, justly praised for almost 40 years. Any closely attentive reader who knows schoolroom arithmetic and the use of exponents is prepared for these pages, and few outside the profession will find the material too familiar.

Consider chapter four. Four is the first integer that is also the square of a smaller integer, its most familiar property. That mathematician of Padua, Galileo Galilei himself, realized first of all what he let his character Salviatus state: the squares of all the integers, a list headed by 4, was long known to be an unending list. Yet that list of squares is demonstrably neither shorter nor longer than the list of all the integers. Salviatus argued in modern vein that therefore "the Attributes of Equality, Majority, and Minority have no place in Infinities, but only in terminate quantities." Any infinite set can be matched one to one with a part of itself! Aleph-zero, the

infinity of integers, would be followed by aleph-one, which Cantor expected to be the denser infinity, all the points in a continuous line segment. That plausible conclusion is now itself moot, demonstrated in the 1960s as genuinely undecidable from the axioms of arithmetic.

FEARFUL SYMMETRY: IS GOD A GEOMETER? by Ian Stewart and Martin Golubitsky. Blackwell Publishers, 1992 ($24.95).

A modern exercise in natural philosophy, this lively, well-illustrated book is devoted to the currently important topic of symmetry-breaking. Symmetry-breaking is the insight that a deep-lying symmetry of cause can somehow "be lost between cause and effect." The authors pursue the issue from an Edgerton drop that splashes into a transient, beaded crown past crystals and pseudocrystals to patterned fluid flow, tigerskins à la Turing, and some deeper glimpses of the patterns that arise within disorder. There is plenty of readable, provocative text here, and the topics are so often raised in public discussions that the book is a genuine help to readers, at high school mathematics level and beyond.

The senior of these two authors is well known to readers of this journal; his Texas collaborator is an investigator of pattern within disorder. They write with flair, although they have undergone occupational exposure to whimsy, whose terminal cuteness they have survived better than some of the authors they cite. Their conclusions are moderate. What we see as symmetry in matter is almost always approximate, for we cannot see the granular, atomic detail in what looks to us like a neatly spherical drop. It is very often the amplification of that fine structure that "breaks" strongly what was only an approximate symmetry in the first place. If God is a geometer, She is no narrowly rigorous group theorist but approximates wonderfully, down there in the details. Scale is what matters.

THE KIDS' CODE AND CIPHER BOOK, by Nancy Garden. Linnet Books, 1991 ($17.50).

The experiences of Captain Snow, a cipher expert under sail, and his missing son, Samuel, doing their best to avoid pirates, offer a fictional frame for codes and ciphers. The strength of the book is its admirable practicality for kids. Its first, very simple ciphers—nothing more than backwards spelling—demonstrate the sensible schemes of adding nulls to text, suppressing cases and word spaces and other devices to confuse the unwanted reader. Quickly the scrambling becomes more complicated: twisted paths drawn on a graph-paper plain text, for instance. Then we see the classical transpositions, Caesar's shifted alphabets, checkerboard ciphers and more, all in service as coding and decoding practice.

The menu changes from letter hash toward writing with strange symbols: Pigpen corners and dots, Morse, Ogham, Runes, musical notes, Holmes' Dancing Men. At last the forms combine into one tough message. On to simple code and cipher machines a kid can build and use. Start with the telephone dial, proceed to stencillike grills and to movable St. Cyr alphabet tapes. There are strong hints on organizing effective and patient work, answers for problems and a good annotated list of other books. This book for young enthusiasts is meant as a guide to ingenious action, not as a history of adult codes nor an exploration of much mathematics of substitutions. Oh, yes, very easily: TA TSAL EHT ETARIP EGRUOCS SI DEDNE.

Earth

THE HISTORY OF EARTH: AN ILLUSTRATED CHRONICLE OF AN EVOLVING PLANET, by William K. Hartmann. Paintings by Ron Miller and by the author. Workman Publishing Company, 1991 ($35; paperbound, $19.95).

Once the classical view held wild, unreasoning nature as dangerous. Today's romanticism sees nature as good, civilization bad. To respond well to what we know of our earth, we need both stances, evoked here in an original, colorful and personal story, full of twists and turns, honestly argued from the best current evidence, with graphs and some numbers, but not formulas.

The visually documented narrative treats five billion years past and forecasts six billion years more to come. It opens with the formation of Earth and moon during the merging collision of a passing protoplanet with our proto-Earth (a view the author pioneered) and extends to the end game of Earth and sun. First Earth will survive a slow, chancy encounter of our Milky Way with its neighbor galaxy, then either Earth will endure full engulfment within a red and swollen sun or pass into a frozen epoch of eventless rounds around a dwarfed and cheerless one. Our fragile posterity will perhaps have transcended their earthborn origins or will come to the end of consciousness, while in some yet unknown star-stream others will likely live out "lives not altogether unfamiliar" to our kind.

Fourteen paintings record the amazing spectacle when last the big asteroids fell, the sea boiled, the forests burned and meek mammals inherited the damaged earth. One scene beforehand shows the quiet Cretaceous forest; another ends the drama at an overlook above the quiet, sea-flooded crater 1,000 years after impact. This is an audacious, imaginative, instructive work, so compact of what is and of what might be that no one today can well sift them apart. Readers who enjoy following scientific argument will not soon put the volume down.

HUDSON RIVER: AN ADVENTURE FROM THE MOUNTAINS TO THE SEA, by Peter Lourie. Illustrated with full-color photographs. Caroline House, 1992 (Distributed by St. Martin's Press) ($15.95).

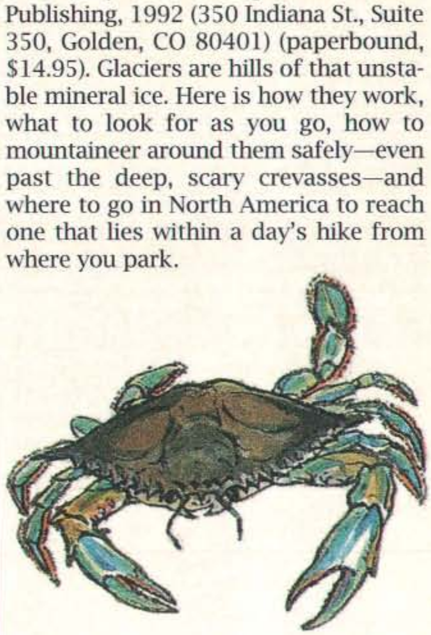
The Hudson River begins as a mountain cascade too small and boulder-filled to carry a canoe out of little Lake Tear of the Clouds, where the river rises. Soon enough, canoeists must prepare for whitewater battles. Rapids crossed, portages behind, the author comes down the wild river to a river of human activity: nine power dams to shunt, then the great hospitable locks of the Champlain Canal. By this point the river carries big ships and sorry pollution. Still far upstream, you feel the sea tides. Now it is the moon's and not the earth's pull that fills the traveler's thoughts; he must plan by the clock to make headway. Down with the current he paddles each day, past the brow of Storm King, past a mock castle, past great bridges, until after three weeks on the river he sways in the salt sea at Battery Park, around him "mountains...made of glass and steel," 315 miles below swift Feldspar Outlet. This tale of flow manages to give its readers (eight years to 80) a real intimacy of place and the heady sense of far voyaging at one and the same time.

Field Guides

In these books, print nourishes activity. A field guide shows a variety of concrete examples held within some ordering framework that can support firsthand extension of experience. This year we reaped a wide diversity of guides for young and old. Let your interests lead your choice. Any one of them belongs on some handy shelf, ready to stimulate an expedition in its season or to enlarge

into reality some enticing hint from conversations or the news.

GLACIERS OF NORTH AMERICA: A FIELD GUIDE, by Sue A. Ferguson. Fulcrum Publishing, 1992 (350 Indiana St., Suite 350, Golden, CO 80401) (paperbound, $14.95). Glaciers are hills of that unstable mineral ice. Here is how they work, what to look for as you go, how to mountaineer around them safely—even past the deep, scary crevasses—and where to go in North America to reach one that lies within a day's hike from where you park.



PETERSON FIRST GUIDE TO SEASHORES, by John C. Kricher. Illustrated by Gordon Morrison. Houghton Mifflin Company, 1992 (paperbound, $4.95). An inexpensive, abridged, hand-sized opener to recognizing some of the animals, plants, birds, fish, shells, worms and trees on all the shores of the U.S. The deft paintings are marked to emphasize visual clues. This is one of a dozen new primers that derive from a classic series of comprehensive guides.

THE AURORA WATCHER'S HANDBOOK, by Neil Davis. University of Alaska Press, 1992 ($35; paperbound, $20). The forms and colors of the northern lights (the southern lights in the Antarctic are not ignored) are represented by striking photographs from the earth and from space. Up-to-date physical explanations begin at a simple level and go beyond. South of the Canadian border you need patience to view any aurora—let alone to expect to hear one.

POISON IVY, POISON OAK, POISON SUMAC AND THEIR RELATIVES, by Edward Frankel. Drawings by Anthony Salazar. Boxwood Press, 1991 (183 Ocean View Blvd., Pacific Grove, CA 93950) ($9.95). A stable, oily resin in all plants of this genus is a complex "oil of dermatitis," inducing allergic reaction by contact in three Americans out of four. Practical for the outdoors, the fascinating account here often surprises. Even delicious mangoes secrete related irritating compounds; wash and peel them well if you fear you are in harm's way.

FURTIVE FAUNA: A FIELD GUIDE TO THE CREATURES WHO LIVE ON YOU, by Roger M. Knutson. Penguin Books, 1992 (paperbound, $8). Mosquitoes and chiggers are only visitors, fleas only intrusive neighbors. But tiny follicle mites are true long-term residents, grossly if harmlessly living, eating and breeding right in your face. Drawings and details, but no clear clues to finding mites in your eyebrows.

WHERE TO FIND DINOSAURS TODAY, by Daniel and Susan Cohen. Puffin Unicorn, 1992 (paperbound, $6.99). They are widespread: two huge dino sculptures in concrete, twice as big as life, stand next to an all-night restaurant near Palm Springs. Seriously scientific to plain silly, what you can find ranges from toys and robots to real footprints and half-buried bones, up to big assembled skeletons and well-restored monsters. Two sharp eyewitnesses point out where to stalk them all, listing museums small and great, specialty stores and striking open-air sites all over the U.S. and Canada.

FOSSIL COLLECTING IN THE MID-ATLANTIC STATES, WITH LOCALITIES, COLLECTING TIPS, AND ILLUSTRATIONS OF MORE THAN 450 FOSSIL SPECIMENS, by Jasper Burns. Johns Hopkins University Press, 1991 (paperbound, $18.95). Exciting hunts for real fossils are not restricted to broad vistas in Montana. This author-collector draws the very roadside cut (and explains where to park) where he reliably finds trilobites. It is on busy, four-lane Route 50, near the "Hebron Baptist Church" outside Winchester, Va., one of 46 sites in this regional guide. If trilobites and Virginia seem far away from you, seek out the apropos regional guides in museums, bookstores and survey offices anywhere.

THE BODY ATLAS, by Mark Crocker. Oxford University Press, 1991 ($16.95). From individual cells to a dozen vital systems, skin to skeleton, this book opens the natural world of the body seen over time to any middle grader. All drawings are large, colorful, effective and candid. The opening metaphor, a map of the human body as a landscape complete with roads, offers an unusually inviting start.

THE PRACTICAL ENTOMOLOGIST, by Rick Imes. Simon & Schuster/Fireside, 1992 (paperbound, $15). Tells not only where bugs, flies, dragonflies, wasps, moths, aphids, grasshoppers and bee-

tles abound, but how to recognize and to keep them, alive or mounted.

Labs

THE EXPLORATORIUM SCIENCE SNACKBOOK, by the Exploratorium Teacher Institute. The Exploratorium, 1991 (3601 Lyon St., San Francisco, CA 94123) (paperbound, $19.95). THE SCIENCE BOOK OF THE SENSES, by Neil Ardley. Harcourt Brace Jovanovich, 1992 ($9.95).

These "snacks" are 100-plus simple, safe and striking demonstrations or activities in physical science and in human perception. Each is given a page or two of clear description, without jargon but with explanations, photographs, diagrams and full details for forgiving and inexpensive assembly and construction.

Such admirable practicality reflects the process of making this unique book. The San Francisco Exploratorium, a pioneering hands-on museum of science, art and perception, holds in its public space about 650 exhibits. With that treasury as their inspiration, group after group of science teachers in local grade and high schools were invited to participate in the joint task of preparing and trying out modest activities in the same spirit as those on the museum floor. Theirs would be "snacks," simpler and cheaper, to be built not by artists, engineers and craftsmen but by kids and teachers on their own. After some years, success was brilliant, and this communication is genuine: these are recipes you can both understand and make work. Many of the devices will last for years.

What's here? A famous scheme that brings two people face to face, a square foot of half-silvered mirror (actually, a clever cheap equivalent) set between them. Each one can vary the light intensity on his or her face. Identities soon merge. Or make an electroscope out of plastic tape, a pinhole magnifier, bubbles that float stably in the midst of "air," an endlessly overhanging stack of wooden slats, a surprising photographed face to be viewed upside-down... and a hundred more. Energetic teachers and their students need to walk through this network of paths into science; the effort will generate valuable gifts for friends and schools.

Like the *Snackbook,* the small book of experiments using the senses is a workable pathway to simple scientific experience. It is very much simpler, intended for children up to 10 to use at home or in free time, without involvement of teachers or parents. The sup-

plies are simpler still, household items like scissors, fruit juice and kitchen glasses, rubber bands, markers, paper. The activities are those of perception: braillelike reading in the dark, the letters made with pins, illusions of touch and of hot and cold, the difficulty of sound-source location using only one ear, and half a dozen more.

VISUAL MAGIC, by David Thomson. 3-D glasses included. Dial Books/Penguin Books, 1991 ($14.95).

Thirty optical illusions are collected here. They are so well known a set that it is hardly necessary to do more than name them: the spotted dog hidden among his spots, the figure-ground ambiguity of face and vase, an Escher house of unending steps, a number of mystifying prongs, a nonspiral spiral, a bird to be caged, random stereo dots, a red/green stereo image, a big, shimmery white grid on very strong black and some familiar op art.

It is therefore not novelty that makes this book attractive; rather it is the admirable presentation of many well-known experiences essential for young people, using good color, paper, lines and scale. A few of the offerings are in fact out-of-the-ordinary, a painting by Giuseppe Arcimboldo displayed upside-down, a powerful piece of wavy-zebra op art that at a distance discloses a face and instructions for the Pulfrich moving-pendulum illusion and for viewing your own retina, neither perhaps easy to carry out, both of them worth much effort.

KIDS SHENANIGANS: GREAT THINGS TO DO THAT MOM AND DAD WILL JUST BARELY APPROVE OF, by the editors of Klutz Press. Cartoons by H. B. Lewis, instructional art by Sara Boore. Klutz Press, 1992 (2121 Staunton Ct., Palo Alto, CA 94306) (spiralbound, $13.95, with Whoopie Cushion kit).

"What kids and teens *really* want to do, and what their parents are prepared to let them do, are almost entirely not the same things." So these savants conclude from the revealing testimony of many "detention-room veterans with scary...reputations." There remains a small region of overlap, definitely inside the Real Fun Zone yet still within the Parent Approval Line ("or at least close" to it).

The shenanigans herein number about 30. We will suggest them by sampling about half. Open calmly with The Best Paper Airplane: all but respectable, and certainly exacting, with 13 sequential drawings. Then there is the amazing Tennis Ball Launcher and a careful drill in the piercing sounds of Hand Whistling, both of them with high redeeming physics value. Flinch and Stink Eye, the applied genetics of Body Weirdnesses, and a few others are easily tolerable because they naturally engage the kids themselves: divide and rule.

But this is no time for complacency. Dead Finger, a real finger presented bloodied in a box; Hand Jive, half a dozen interminable and insulting handshakes; Upside-Down Water, a present danger of flood on the dining room table; Fake a Sneeze (another wet one); and the (ugh) Whoopie Cushion and its kit (you hardly want to know more) clearly reach the intensity labeled Gross on the Parental Beaufort Scale and ought seldom to occur on land.

The One Finger Grown-Up Hoist, four or five kids jointly lifting a person by their forefingers; Hanging Spoons that depend from a kid's spit-wetted nose; the Parent and the Egg, an egg placed gently into the ready fingers that guess-who cooperatively (and foolishly) holds out through the front-door hinge opening; How to Sneak Around, a tour de force in five tight text pages; Grown-Up Stumpers, neat trick questions for people gone beyond the age of maximum brain size—these are grim weapons indeed, but within the tolerant rules of engagement between generations.

After all that, how could the undescribed half hold any surprises? (It does.) Cultivated parents with sharp memories of adolescence and kids who spend their own money well should make this book a popular and a critical triumph, although some may cite it as grounds for seeking disaster relief. The guilty parties include people as inventive as Martin Gardner and Klutz-in-Chief John Cassidy, whose autobiographies are perhaps just legible here between the lines.

ANNUAL INDEX
1992

AUTHORS

ARTICLES

Reprints may be ordered from Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111, or fax (212) 355-0408.

THE AMATEUR SCIENTIST

MATHEMATICAL RECREATIONS

ESSAY by Anne Eisenberg

Imperial English: The Language of Science?

Werner Heisenberg learned Latin, Greek and French when he was a gymnasium student in Munich. Later, when he worked in Copenhagen, he tackled English and Danish, using mealtimes as his language lab: English conversation during breakfast; Danish read aloud from the newspaper by his landlady afterward.

This is not the kind of anecdote we associate with today's science majors in the U.S., that resolutely monolingual lot. Science students here are rarely to be found in a school language lab, much less a spontaneous one, and when they do speak another language, it is usually because of family background, not classroom instruction. Then they graduate, attend a conference with colleagues from other countries and discover the international hallmark of U.S. science: linguistic incompetence.

We are the people who can no longer be bothered to learn another language. To be sure, we really haven't had to since the 1960s, for in the years since World War II English has gradually but inexorably become the lingua franca of science. Today it is the universal currency of international publications as well as of meetings. Those of us who need to keep up with, say, *Angewandte Chemie* need not worry about mastering German; we can leave it to the journal's staff, whose English is no doubt immaculate, to provide us with a convenient international edition published, of course, in English.

It wasn't always this way. For the 200 years before World War II, most scientific work was reported in German, French or English, in that order of importance. People who wanted to keep up with a specialization had to learn the dominant language of the field. For example, scientists who wished to understand quantum mechanics in the 1920s had to learn German. Sir Nevill Mott comments, "Apart from Dirac, I don't think anyone in Cambridge understood [quantum mechanics] very well; there were no lectures on it, and so the only thing to do was to learn German and read the original papers, particularly those of Schrödinger and Born's *Wellenmechanik der Stossvorgänge* ['Wave Mechanics of Collision Processes']."

German, French and English were the customary languages of meetings, too. At Niels Bohr's institute in Copenhagen, for example, John A. Wheeler recalls that most seminars were held in German, occasionally in English. Bohr, who spoke English and German with equal ease, fluctuated between them, adding Danish as counterpoint. No one had to learn French, though, for Bohr's knowledge of it was limited. "I have it from an eyewitness," Abraham Pais writes, "that he once greeted the French ambassador to Denmark with a cordial *aujourd'hui.*"

After World War II, the linguistic balance of power shifted. The U.S. economy boomed, and science grew rapidly as vast federal expenditures, often fueled by the cold war, poured into research and development. U.S. scientists flocked to conferences, bringing their language with them; U.S. scientific publications burgeoned, and their huge readerships made them highly desirable to scientists throughout the world who realized English was a medium through which they could be widely read and cited.

With technical dominance came the beginning of linguistic dominance, first in Europe, then globally. Only the French and the Soviets put up a spirited resistance. At one international conference when de Gaulle was still in power, for instance, a member of the French contingent began reporting in French and then, sensing that many of the important U.S. scientists in the audience did not understand him, switched to English. Then he watched as all of his French colleagues rose as a group and exited. The Soviets, too, did what they could to fight the English monopoly, providing expensive simultaneous technical translations and bilingual commentaries or even resorting to French as the lesser evil—anything to avoid the language of the enemy.

Today in the former U.S.S.R., linguistic opposition has dissolved with the union. Even the French, who fiercely cherish their language, have accepted the practicality of English for publishing the proceedings of international meetings: the 12th Colloquium on High Resolution Molecular Spectroscopy was held last year in Dijon, but the only speech in French during the five days of the meeting was the mayor's welcoming address.

The rest of the world's scientists, too, have fallen into step. English was already in place in India, Nigeria and many other countries where it had been left behind by the British, to be widely adopted as a practical second language that united diverse populations. The Japanese readily inserted the language of the victor into their children's school programs; Korean and Chinese scientists were delighted to take up membership in the English-speaking club.

English is indeed the new Latin. It has become a successor to the scholarly language once so powerful that Christian Huygens delayed publishing *Traité de la Lumière* for 12 years in hopes of translating it into Latin so as "to obtain greater attention to the thing." And there is a second way that English may parallel Latin. Latin outlived the Roman Empire, surviving long after the government that spread it through the world had vanished. So may the international use of English outlast U.S. scientific dominion. The ascent of English, after all, had little to do with any inherent linguistic virtues. True, English has an unusually rich vocabulary; instead of resisting new terms, we welcome them, particularly in science and technology—*les anglicismes* have conquered the world. But it was scientific leadership, not a flexible lexicon, that sparked the diffusion of English. Many now say this leadership is faltering. Consider, for instance, last year's top holders of new U.S. patents: Toshiba, Mitsubishi and Hitachi.

This year, although English continues its reign, small changes are in the wind. For example, more than 860 Japanese language programs are running in U.S. schools, and there is even an occasional undergraduate science department promoting German. Who knows, the students enrolling in these foreign language classes might even learn a bit more about English, or, to put it in Goethe's words, *Wer fremde Sprachen nicht kennt, weiss nichts von seiner eigenen.*

ANNE EISENBERG is a professor at Polytechnic University in Brooklyn, where a very popular Japanese language course was recently begun.